

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:39:00 ; Search time 36.5039 Seconds
(without alignments)
2747.774 Million cell updates/sec

Title: US-09-597-796C-4

Perfect score: 1802

Sequence: 1 MNSRRSRSLRWSLLSLAA.....QTKSGTGTGNTVLAEGPPA 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 segs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1802	100.0	355	2	AAW32367
2	1802	100.0	355	2	AAW32435
3	1802	100.0	355	2	AAW64307
4	1802	100.0	355	2	AAW81670
5	1802	100.0	355	2	AAW38972
6	1802	100.0	355	2	AAW39109
7	1802	100.0	355	4	AAU01890
8	1802	100.0	355	5	AAE29701
9	1802	100.0	355	5	AAE17565
10	1792	99.4	355	2	AAW05000
11	1792	99.4	355	4	AAO21137
12	1792	99.4	355	4	AAW81110
13	1792	99.4	379	2	AAW04830
14	1634	90.7	330	5	AAE29702
15	1634	90.7	330	5	AAE17566
16	1631	90.5	330	5	AAE29703
17	1631	90.5	330	5	AAE17567
18	1631	90.5	330	7	ADA26372
19	1631	90.5	723	7	ADA26354
20	1631	90.5	1010	7	ADA26356
21	1627	90.3	330	7	ADA26371
22	999.5	55.5	1016	7	ADA26370
23	992	55.0	1154	7	ADA26368
24	990.5	55.0	825	7	ADA26366
25	989.5	54.9	813	7	ADA26367

26	988	54.8	1022	7	ADA26369	Ada26369 Mycobacte
27	985	54.7	195	5	AAE29704	Aae29704 Mycobacte
28	985	54.7	195	5	AAE17568	Aae17568 Mycobacte
29	983.5	54.6	875	7	ADA26365	Ada26365 Mycobacte
30	983	54.6	596	2	AAW32070	Aaw32070 Mycobacte
31	983	54.6	596	5	AAE29710	Aae29710 Mycobacte
32	983	54.6	596	5	AAE17574	Aae17574 Mycobacte
33	983	54.6	599	5	AAU74599	Aau74599 Antigenic
34	983	54.6	729	4	AAO22142	Aao22142 Ra12-H9-3
35	983	54.6	729	5	AAE29708	Aae29708 Mycobacte
36	983	54.6	729	5	AAE17572	Aae17572 Mycobacte
37	983	54.6	729	7	ADA26373	Ada26373 Mycobacte
38	983	54.6	930	5	AAE29731	Aae29731 Mycobacte
39	983	54.6	930	7	ADA26364	Ada26364 Mycobacte
40	980	54.4	729	5	AAE29709	Aae29709 Mycobacte
41	980	54.4	729	7	AAE17573	Aae17573 Mycobacte
42	980	54.4	729	7	ADA26374	Ada26374 Mycobacte
43	978	54.3	726	5	AAU74588	Aau74588 Antigenic
44	978	54.3	729	2	AAW32059	Aaw32059 Mycobacte
45	676	37.5	132	2	AAW32354	Aaw32354 Mycobacte

ALIGNMENTS

RESULT 1

AAW32367
ID AAW32367 standard; protein; 355 AA.

XX
AC AAW32367;
XX
DT 13-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TBra35.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9709429-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US014675.
XX
PR 01-SEP-1995; 95US-00523435.
PR 22-SEP-1995; 95US-00532136.
PR 22-MAR-1996; 96US-00620280.
PR 05-JUN-1996; 96US-00658800.
PR 12-JUL-1996; 96US-00680573.
PA (CORI-) CORIXA CORP.
XX
XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
XX Vedvick TH, Twardzik DR;
XX WPI; 1997-192904/17.
XX N-PSDB; AAT91414.
PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
PT useful for diagnosis of M. tuberculosis infection.
PS Example 3; Page 124-126; 190pp; English.
XX
CC A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
CC variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a M. tuberculosis antigen,
CC TBra35. The immunogenic polypeptide can be used to diagnose
CC M. tuberculosis infection by forming complexes with specific antibodies in
CC the sample. Fragments of DNA encoding the immunogenic polypeptide can be
CC used as diagnostic primers or probes and agents that bind to the antigen,
CC

CC especially monoclonal antibodies or equivalent polyclonal antibodies, are
 CC also used for diagnosis

XX Sequence 355 AA;

Query Match 100.0%; Score 1802; DB 2; Length 355;
 Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSLAAVGLGLATAPAAQAPALSDREAPALPLDPSAMVAQA 60
 DB 1 MNSRRSLRWSLLSLAAVGLGLATAPAAQAPALSDREAPALPLDPSAMVAQA 60
 QY 61 PQVNTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSGSGGTGVVDVVG 120
 DB 61 PQVNTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSGSGGTGVVDVVG 120
 QY 121 YRTQDVAVLQLRGAGLPSAIGGVAVGEFVWAMGNSGGGTPRAVPGRVVAGQTV 180
 DB 121 YRTQDVAVLQLRGAGLPSAIGGVAVGEFVWAMGNSGGGTPRAVPGRVVAGQTV 180
 QY 181 QASDSLTAETLNGLIQFDAAIQPDSCGPPVNGLVGQVGNNTAASDNFQLSQGGQGA 240
 DB 181 QASDSLTAETLNGLIQFDAAIQPDSCGPPVNGLVGQVGNNTAASDNFQLSQGGQGA 240
 QY 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGARVORVVGSAAPASLGIST 300
 DB 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGARVORVVGSAAPASLGIST 300
 QY 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNMQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNMQTKSGGTRTGNVTLAEGPPA 355

RESULT 2

AAW32435
 ID AAW32435 standard; protein; 355 AA.

XX AAW32435;

DT 08-JAN-1998 (first entry)

DE Mycobacterium tuberculosis antigen TBra35.

KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
 KW skin testing; M.tuberculosis.

XX Mycobacterium tuberculosis.

OS WO9709428-A2.

PN 13-MAR-1997.

XX 30-AUG-1996; 96WO-US014674.

XX 01-SEP-1995; 95US-00523436.

PR 22-SEP-1995; 95US-00533634.

PR 22-MAR-1996; 96US-00620874.

PR 05-JUN-1996; 96US-00659683.

PR 12-JUL-1996; 96US-00680574.

XX (CORI-) CORIXA CORP.

PA Reed SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;

PI Vedwick TH, Twardzik DR;

XX WPI; 1997-192903/17.

DR N-PSDB; AAR91477.

XX New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
 PT useful in vaccines for prevention or treatment of tuberculosis, also for
 PT diagnosis.

PS Example 3; Page 114-116; 168pp; English.

XX A new immunogenic polypeptide has been developed comprising an
 CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
 CC variant differing only in conservative substitutions and/or
 CC modifications). The present sequence represents a M.tuberculosis antigen,
 CC TBra35. The immunogenic protein, and fusion proteins containing one or
 CC more of the proteins or one of the proteins plus ESAT-6, are useful in
 CC vaccines, preferably when formulated with a non-specific adjuvant, to
 CC induce an immune response against M.tuberculosis (for treatment or
 CC prevention)

XX Sequence 355 AA;

Query Match 100.0%; Score 1802; DB 2; Length 355;

Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLSLAAVGLGLATAPAAQAPALSDREAPALPLDPSAMVAQA 60
 DB 1 MNSRRSLRWSLLSLAAVGLGLATAPAAQAPALSDREAPALPLDPSAMVAQA 60
 QY 61 PQVNTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSGSGGTGVVDVVG 120
 DB 61 PQVNTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSGSGGTGVVDVVG 120
 QY 121 YRTQDVAVLQLRGAGLPSAIGGVAVGEFVWAMGNSGGGTPRAVPGRVVAGQTV 180
 DB 121 YRTQDVAVLQLRGAGLPSAIGGVAVGEFVWAMGNSGGGTPRAVPGRVVAGQTV 180
 QY 181 QASDSLTAETLNGLIQFDAAIQPDSCGPPVNGLVGQVGNNTAASDNFQLSQGGQGA 240
 DB 181 QASDSLTAETLNGLIQFDAAIQPDSCGPPVNGLVGQVGNNTAASDNFQLSQGGQGA 240
 QY 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGARVORVVGSAAPASLGIST 300
 DB 241 IPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGARVORVVGSAAPASLGIST 300
 QY 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNMQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNMQTKSGGTRTGNVTLAEGPPA 355

RESULT 3

AAW64307
 ID AAW64307 standard; protein; 355 AA.

XX AAW64307;

AC AAW64307;

DT 17-OCT-2003 (revised)

DT 09-NOV-1998 (first entry)

XX Mycobacterium tuberculosis antigen TBra35.

DE Tuberculosis; infection; diagnosis; antigen; TBra35.

XX Mycobacterium tuberculosis; strain H37Ra.

XX WO9816645-A2.

XX 23-APR-1998.

XX 07-OCT-1997; 97WO-US018214.

XX 11-OCT-1996; 96US-00729622.

PR 13-MAR-1997; 97US-00818111.

XX (CORI-) CORIXA CORP.

XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;

PI Vedwick TS, Twardzik DR, Lodes MJ;

XX WPI; 1998-251292/22.

DR N-PSDB; AAV44355.
XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
PT develop products for the detection of M. tuberculosis infection and
PT diagnosis of tuberculosis.
XX
XX
XX Example 3; Page 115-116; 250pp; English.
PS
XX This polypeptide comprises Mycobacterium tuberculosis soluble antigen
CC Tbra35. It is encoded by a DNA sequence (see AAV44355) isolated from a M.
CC tuberculosis strain H37Ra expression library with rabbit anti-sera raised
CC against M. tuberculosis supernatant. No significant homology was found
CC between Tbra35 and Genbank database sequences. The invention relates to
CC compositions and methods for diagnosing tuberculosis. It provides
CC polypeptides (see AAV4291-W4379) comprising an antigenic portion of a
CC soluble M. tuberculosis antigen, or an immunogenic portion of an M.
CC tuberculosis antigen, as well as DNA sequences encoding such
CC polypeptides, recombinant expression vectors and transformed or
CC transfected host cells. Also claimed are methods and diagnostic kits for
CC detecting M. tuberculosis infection in a patient using the above
CC polypeptides, antibodies or oligonucleotide probes and primers, for the
CC diagnosis of tuberculosis. (Updated on 17-OCT-2003 to standardise OS
CC field)
XX
XX SQ Sequence 355 AA;
Query Match 100.0%; Score 1802; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 7.2e-126;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MSNSRRSLRWLLSVLAAGVGLGLATAPACAPALSDQRFADFPALPLDPSAMVAQVA 60
Db 1 MSNSRRSLRWLLSVLAAGVGLGLATAPACAPALSDQRFADFPALPLDPSAMVAQVA 60
Qy 61 PQVNINTKLYNNVAGAGTGVDPNGVLTNNHVIAGATDINAFVSGGQTYGVDDVG 120
Db 61 PQVNINTKLYNNVAGAGTGVDPNGVLTNNHVIAGATDINAFVSGGQTYGVDDVG 120
Qy 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVAMNGSGGGTTPRAVGRVVALGQTV 180
Db 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVAMNGSGGGTTPRAVGRVVALGQTV 180
Qy 181 QASDLSLTGAETLNGLIQFDAAIQDSDGGPVVNGLGQVGMNTAASDNFQLSGGGQGFA 240
Db 181 QASDLSLTGAETLNGLIQFDAAIQDSDGGPVVNGLGQVGMNTAASDNFQLSGGGQGFA 240
Qy 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGARVORVVGSAAPASLGIST 300
Db 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGARVORVVGSAAPASLGIST 300
Qy 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
Db 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
RESULT 4
AAW81670
ID AAW81670 standard; protein; 355 AA.
XX
XX AAW81670;
XX
XX 27-JAN-1999 (first entry)
XX
XX M. tuberculosis immunogenic polypeptide Tbra35.
XX
XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
KW vaccine; pharmaceutical; infection; diagnosis.
XX
XX Mycobacterium tuberculosis.
XX
XX WO9816646-A2.
PN
XX 23-APR-1998.
PD

XX
PF 07-OCT-1997; 97WO-US018293.
XX
PR 11-OCT-1996; 96US-00730510.
PR 13-MAR-1997; 97US-00818112.
XX
XX (CORI-) CORIXA CORP.
XX
XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedwick TS, Twardzik DR, Lodes MJ;
XX
XX WPI: 1998-261042/23.
DR N-PSDB; AAV64463.
XX
XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
PT develop products for the detection of M. tuberculosis infection and for
PT diagnosis, treatment and prevention of tuberculosis.
XX
XX Example 3; Page 110-111; 230pp; English.
PS
XX This sequence represents an immunogenic portion of a soluble
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for
CC inducing protective immunity against tuberculosis (TB). This sequence can
CC be formulated into vaccines and/or pharmaceutical compositions for
CC immunising against M. tuberculosis infection or may be used for the
CC diagnosis of tuberculosis
XX
XX SQ Sequence 355 AA;
Query Match 100.0%; Score 1802; DB 2; Length 355;
Best Local Similarity 100.0%; Pred. No. 7.2e-126;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MSNSRRSLRWLLSVLAAGVGLGLATAPACAPALSDQRFADFPALPLDPSAMVAQVA 60
Db 1 MSNSRRSLRWLLSVLAAGVGLGLATAPACAPALSDQRFADFPALPLDPSAMVAQVA 60
Qy 61 PQVNINTKLYNNVAGAGTGVDPNGVLTNNHVIAGATDINAFVSGGQTYGVDDVG 120
Db 61 PQVNINTKLYNNVAGAGTGVDPNGVLTNNHVIAGATDINAFVSGGQTYGVDDVG 120
Qy 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVAMNGSGGGTTPRAVGRVVALGQTV 180
Db 121 YDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVAMNGSGGGTTPRAVGRVVALGQTV 180
Qy 181 QASDLSLTGAETLNGLIQFDAAIQDSDGGPVVNGLGQVGMNTAASDNFQLSGGGQGFA 240
Db 181 QASDLSLTGAETLNGLIQFDAAIQDSDGGPVVNGLGQVGMNTAASDNFQLSGGGQGFA 240
Qy 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGARVORVVGSAAPASLGIST 300
Db 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGARVORVVGSAAPASLGIST 300
Qy 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
Db 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
RESULT 5
AAV38972
ID AAV38972 standard; protein; 355 AA.
XX
XX AAV38972;
XX
XX 05-NOV-1999 (first entry)
XX
XX M. tuberculosis recombinant antigen protein Tbra35.
XX
XX Antigen; diagnosis; detection; infection; antibody; immunisation;
KW vaccine; immunity.
XX
XX Mycobacterium tuberculosis.
XX
XX OS
XX

PN WO9942118-A2.
 XX 26-AUG-1999.
 XX 17-FEB-1999; 99WO-US003265.
 XX 18-FEB-1998; 98US-00024753.
 PR 03-MAY-1998; 98US-00072596.
 XX (CORI-) CORIXA CORP.
 XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MU, Hendrickson RC;
 XX WPI; 1999-527416/44.
 DR N-PSDB; AAZ19053.
 XX New polypeptide comprising antigenic portions of M. tuberculosis.
 XX Example 3; Page 151-153; 323pp; English.
 CC This invention describes novel recombinant antigens and their encoding
 CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 CC polypeptides are useful for detecting M. tuberculosis infection in a
 CC biological sample by detecting antibodies which bind with the
 CC polypeptides, and are useful as vaccines for immunizing against M.
 CC tuberculosis infection. The new detection methods are needed as current
 CC vaccination strategies do not provide 100% immunity
 XX Sequence 355 AA;
 SQ
 Query Match 100.0%; Score 1802; DB 2; Length 355;
 Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MNSRRSRSLRWLLSVLAAGVGLGATAPAAQAPALSDQDRFADFPALPLDPSAMVAQA 60
 DB 1 MNSRRSRSLRWLLSVLAAGVGLGATAPAAQAPALSDQDRFADFPALPLDPSAMVAQA 60
 QY 61 PQVNTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
 DB 61 PQVNTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
 QY 121 YDRTQDVAVLQRGAGGLPSAAIGGVAVGEPVAVMGNSGGQGTTPRAVGRVVALGQTV 180
 DB 121 YDRTQDVAVLQRGAGGLPSAAIGGVAVGEPVAVMGNSGGQGTTPRAVGRVVALGQTV 180
 QY 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVGMNTAASDNFQLSQGGQGA 240
 DB 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVGMNTAASDNFQLSQGGQGA 240
 QY 241 IPIGQAMAIAGIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVQRVVGSPAPASLGIST 300
 DB 241 IPIGQAMAIAGIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVQRVVGSPAPASLGIST 300
 QY 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
 RESULT 6
 AAY39109
 ID AAY39109 standard; protein; 355 AA.
 XX AAY39109;
 XX 05-NOV-1999 (first entry)
 XX M. tuberculosis antigen TBRA35 amino acid sequence.
 XX Mycobacterium tuberculosis; M. tuberculosis; antigen; immunogen;
 KW immunotherapy; diagnosis; immunisation; vaccine; infection;
 KW immune response; skin test.

XX Mycobacterium tuberculosis.
 OS WO9942076-A2.
 PN 26-AUG-1999;
 XX 17-FEB-1999; 99WO-US003268.
 XX 18-FEB-1998; 98US-00025197.
 PR 05-MAY-1998; 98US-00072967.
 XX (CORI-) CORIXA CORP.
 XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
 XX WPI; 1999-527409/44.
 DR N-PSDB; AAZ19265.
 XX New antigens from Mycobacterium tuberculosis useful in diagnostic skin
 XX tests and protective or therapeutic vaccines or compositions.
 XX Example 3; Page 108-109; 299pp; English.
 CC The present invention describes polypeptides comprising an immunogenic
 CC part of a Mycobacterium tuberculosis antigen (Ag). Also described are
 CC vaccines and fusion protein containing M. tuberculosis Ag's. M.
 CC tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
 CC polypeptides fragments, can be used in pharmaceutical compositions or
 CC vaccines to generate a protective or therapeutic immune response to M.
 CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
 CC Ag can induce proliferation of, or cytokine secretion by, T, B or natural
 CC killer cells and/or macrophages in tuberculosis-immune subjects. AAZ19249
 CC to AAZ19460 and AAY39083 to AAY39225 are used in the exemplification of
 CC the present invention
 XX Sequence 355 AA;
 SQ
 Query Match 100.0%; Score 1802; DB 2; Length 355;
 Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MNSRRSRSLRWLLSVLAAGVGLGATAPAAQAPALSDQDRFADFPALPLDPSAMVAQA 60
 DB 1 MNSRRSRSLRWLLSVLAAGVGLGATAPAAQAPALSDQDRFADFPALPLDPSAMVAQA 60
 QY 61 PQVNTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
 DB 61 PQVNTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
 QY 121 YDRTQDVAVLQRGAGGLPSAAIGGVAVGEPVAVMGNSGGQGTTPRAVGRVVALGQTV 180
 DB 121 YDRTQDVAVLQRGAGGLPSAAIGGVAVGEPVAVMGNSGGQGTTPRAVGRVVALGQTV 180
 QY 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVGMNTAASDNFQLSQGGQGA 240
 DB 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVGMNTAASDNFQLSQGGQGA 240
 QY 241 IPIGQAMAIAGIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVQRVVGSPAPASLGIST 300
 DB 241 IPIGQAMAIAGIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVQRVVGSPAPASLGIST 300
 QY 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAEGPPA 355
 RESULT 7
 AAU01890
 ID AAU01890 standard; protein; 355 AA.
 XX AAU01890;
 XX

XX 29-AUG-2001 (first entry)
 XX M. tuberculosis antigen Tba35 (Mtb32A).
 DE Tba35; Mtb32A; antigen; vaccine; tuberculosis; AIDS;
 KW acquired immunodeficiency disease.
 XX Mycobacterium tuberculosis.
 OS
 XX WO200124820-A1.
 XX 12-APR-2001.
 XX 10-OCT-2000; 2000WO-US028095.
 XX 07-OCT-1999; 99US-0156338P.
 PR 07-OCT-1999; 99US-0156425P.
 XX (CORI-) CORIXA CORP.
 XX Skeiky Y, Reed S, Houghton RL, Mcneill PD, Dillon DC, Lodes ML;
 FI WPI; 2001-290576/30.
 XX N-PSDB; AAS03781.
 XX
 XX Vaccinating against Mycobacteria infections in mammals using fusion
 PT proteins comprising combinations of heterologous antigens.
 XX
 XX Disclosure; Page 153-154; 168pp; English.
 XX The sequence represents Mycobacterium tuberculosis Tba35 (also known as
 CC Mtb32A), an M. tuberculosis antigen. Compositions comprising at least 2
 CC heterologous antigens, as a fusion protein, and vectors expressing the
 CC fusion proteins are used as vaccines to prophylactically immunise mammals
 CC (especially humans) against infection by Mycobacteria. The compositions
 CC contain at least 2 heterologous antigens that increase the serological
 CC sensitivity of individuals infected with tuberculosis, a disease
 CC frequently affecting patients with acquired immunodeficiency disease,
 CC AIDS
 XX
 XX Sequence 355 AA;
 XX
 XX Query Match 100.0%; Score 1802; DB 4; Length 355;
 XX Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 XX Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSNRRRLRWLLSVLAAVGLGLATAPAAQAPALSSQDRFADFPALPLDPSAMVAQVA 60
 DB 1 MSNRRRLRWLLSVLAAVGLGLATAPAAQAPALSSQDRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNVNITKLYNNNAVAGTGIVIDPNGVLTNNHVIAGATDINAFSGSGQTYGVDDVVG 120
 DB 61 PQVNVNITKLYNNNAVAGTGIVIDPNGVLTNNHVIAGATDINAFSGSGQTYGVDDVVG 120
 QY 121 YDRTDQVAVLQIRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPRVVALGQTV 180
 DB 121 YDRTDQVAVLQIRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPRVVALGQTV 180
 QY 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSQGGGGA 240
 DB 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSQGGGGA 240
 QY 241 IPIGQAMAIQIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVQRVVGSAASLGIST 300
 DB 241 IPIGQAMAIQIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVQRVVGSAASLGIST 300
 QY 301 GDVITAVDGAIPNSATAMADALNGHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGAIPNSATAMADALNGHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 8

AAE29701
 ID AAE29701 standard; protein; 355 AA.
 XX
 AC AAE29701;
 XX
 DT 27-JAN-2003 (first entry)
 XX
 DE Mycobacterium tuberculosis Mtb32A antigenic protein.
 XX
 KW Vaccine; immunity; diagnostic agent; gene therapy; Mtb32A antigen;
 KW Ra35FL.
 XX
 OS Mycobacterium tuberculosis.
 XX
 XX WO200272792-A2.
 XX 19-SEP-2002.
 XX 13-MAR-2002; 2002WO-US008223.
 PF 13-MAR-2001; 2001US-0275837P.
 PR 13-MAR-2001; 2001US-0275837P.
 XX (CORI-) CORIXA CORP.
 XX
 XX Skeiky Y, Brannon M, Guderian J;
 FI WPI; 2002-759844/82.
 XX N-PSDB; AAD47076.
 DR
 XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 XX tuberculosis.
 XX
 PS Disclosure; Page 78-79; 155pp; English.
 XX The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is M.
 CC tuberculosis Mtb32A antigenic protein. Mtb32A is also referred to as
 CC Ra35FL
 XX
 XX Sequence 355 AA;
 XX
 XX Query Match 100.0%; Score 1802; DB 5; Length 355;
 XX Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 XX Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MSNRRRLRWLLSVLAAVGLGLATAPAAQAPALSSQDRFADFPALPLDPSAMVAQVA 60
 DB 1 MSNRRRLRWLLSVLAAVGLGLATAPAAQAPALSSQDRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNVNITKLYNNNAVAGTGIVIDPNGVLTNNHVIAGATDINAFSGSGQTYGVDDVVG 120
 DB 61 PQVNVNITKLYNNNAVAGTGIVIDPNGVLTNNHVIAGATDINAFSGSGQTYGVDDVVG 120
 QY 121 YDRTDQVAVLQIRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPRVVALGQTV 180
 DB 121 YDRTDQVAVLQIRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPRVVALGQTV 180
 QY 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSQGGGGA 240
 DB 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMGNTAASDNFQLSQGGGGA 240

QY 241 IPIGQAMATAGQIRSGGSGPTVHIGPTAFGLGVVDNNGNGARVORVVGSAAPAASLGIST 300
 Db 241 IPIGQAMATAGQIRSGGSGPTVHIGPTAFGLGVVDNNGNGARVORVVGSAAPAASLGIST 300
 QY 301 GDVITAVDGAPINSATAMADALNGHHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
 Db 301 GDVITAVDGAPINSATAMADALNGHHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 9

AAE17565
 ID AAE17565 standard; protein; 355 AA.

XX AC AAE17565;

DT 22-APR-2002 (first entry)

XX Mycobacterium species MTB32A (Ra35FL) protein.

DE Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB32A; Ra32FL protein.
 XX OS Mycobacterium sp.
 XX FN WO200198460-A2.
 XX PD 27-DEC-2001.
 XX XX 20-JUN-2001; 2001WO-US019959.
 XX PF 20-JUN-2000; 2000US-00597796.
 PR 01-FEB-2001; 2001US-0265737P.
 XX XX (CORI-) CORIXA CORP.
 PA

PI Skeiky Y, Reed S, Alderson M;
 XX WPI; 2002-147798/19.
 DR N-PSDB; AAD28335.
 XX

PT Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 PS Claim 70; Page 94-95; 136pp; English.

XX The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB32A (Ra32FL) protein
 XX Sequence 355 AA;
 SQ

Query Match 100.0%; Score 1802; DB 5; Length 355;
 Best Local Similarity 100.0%; Pred. No. 7.2e-126;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNSRRSLRWSWLLSVLAAVGLGLATAPAAQAAPPALSQDRFADFFALPLDPSAMVAQVA 60
 Db 1 MSNSRRSLRWSWLLSVLAAVGLGLATAPAAQAAPPALSQDRFADFFALPLDPSAMVAQVA 60
 QY 61 POVNVINTKLGYNNAVAGAGTGIVIDNGVLTNNHVIAGATDINAFSVSGQTYGVDVVG 120
 Db 61 POVNVINTKLGYNNAVAGAGTGIVIDNGVLTNNHVIAGATDINAFSVSGQTYGVDVVG 120
 QY 121 YDRTQDVAVLQRCAGCLPSAAIGGGVAVGCEPVVAMNGSGGGTTPRAYPGRVVALGQTV 180
 Db 121 YDRTQDVAVLQRCAGCLPSAAIGGGVAVGCEPVVAMNGSGGGTTPRAYPGRVVALGQTV 180
 QY 181 QASDSLGTABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAASDNFQLSQGGQGFA 240
 Db 181 QASDSLGTABETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGWNTAASDNFQLSQGGQGFA 240
 QY 241 IPIGQAMATAGQIRSGGSGPTVHIGPTAFGLGVVDNNGNGARVORVVGSAAPAASLGIST 300
 Db 241 IPIGQAMATAGQIRSGGSGPTVHIGPTAFGLGVVDNNGNGARVORVVGSAAPAASLGIST 300
 QY 301 GDVITAVDGAPINSATAMADALNGHHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
 Db 301 GDVITAVDGAPINSATAMADALNGHHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 10

AAV05000

ID AAV05000 standard; protein; 355 AA.

XX AC AAV05000;

DT 06-JUL-1999 (first entry)

XX DE Mycobacterium species protein sequence 50D.

XX KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection.
 XX OS Mycobacterium sp.
 XX DN WO9909186-A2.
 XX PD 25-FEB-1999.
 XX PF 14-AUG-1998; 98WO-FR001813.
 XX PR 14-AUG-1997; 97FR-00010404.
 PR 11-SEP-1997; 97FR-00011325.
 XX PA (INSP) INST PASTEUR.
 XX

PI Gicquel B, Portnoie D, Lim E, Pellicic V, Guigueno A;
 PI Coguet De La Salmoniere Y;
 XX WPI; 1999-181045/15.
 DR N-PSDB; AAX34251.
 XX

PT Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in infection-associated
 PT protein expression.
 XX

PS Claim 32; Fig 50D; 309pp; French.

XX Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
 CC from various Mycobacterium species microorganisms. The encoding
 CC nucleotide sequences can be used as primers and probes for methods for
 CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection

XX Sequence 355 AA;

Query Match 99.4%; Score 1792; DB 2; Length 355;

Best Local Similarity 99.4%; Pred. No. 4e-125;
Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNSRRSLRWLLSVLAAGVGLGATAPAAQAPALPALSQDRFADFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWLLSVLAAGVGLGATAPAAQAPALPALSQDRFADFPALPLDPSAMVAQV 60
QY 61 PQVNTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDVG 120
DB 61 PQVNTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDVG 120
QY 121 YRTQDVAVLQIRGAGLPSAAIGGVAVGEPVVMGNSGGGTPRAVGRVVALGQTV 180
DB 121 YRTQDVAVLQIRGAGLPSAAIGGVAVGEPVVMGNSGGGTPRAVGRVVALGQTV 180
QY 181 QASDSLTCGAETLNGLIQFDAAIQGDSGGPVVNGLGQVGMNTAASDNFOLSGGGGPA 240
DB 181 QASDSLTCGAETLNGLIQFDAAIQGDSGGPVVNGLGQVGMNTAASDNFOLSGGGGPA 240
QY 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVDNNGGARVQVVGSAPAASLGIST 300
DB 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVDNNGGARVQVVGSAPAASLGIST 300
QY 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAGGPPA 355
DB 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAGGPPA 355

RESULT 11
AAO22137
ID AAO22137 standard; protein; 355 AA.
XX AAO22137;
XX AC AAO22137;
XX AC AAO22137;
DT 03-OCT-2002 (first entry)
XX Mycobacterium tuberculosis MTB32A protein.
XX Ral2; serine protease antigen; MTB32A; Mycobacterium tuberculosis;
XX vaccine; eukaryotic; prokaryotic origin; serum antibody; fusion protein;
XX immunogen; cytokine.
XX OS Mycobacterium tuberculosis.
XX PN WO200125401-A2.
XX PD 12-APR-2001.
XX PF 06-OCT-2000; 2000WO-US027652.
XX PR 07-OCT-1999; 99US-0158585P.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky Y, Guderian J;
XX PR WPI; 2001-266299/27.
XX DR N-PSDB; AAL40768.
XX PT Recombinant nucleic acid molecule for producing high yield expression of
XX desired fusion polypeptides, encodes fusion polypeptide comprising
XX Mycobacterium tuberculosis coding sequence and heterologous polypeptide.
XX PS Disclosure; Fig 1; 39pp; English.
XX CC The invention relates to a recombinant nucleic acid molecule encoding a
XX fusion polypeptide, comprising a polynucleotide sequence of Ral2, a 14
XX kDa C-terminal fragment of serine protease antigen MTB32A of
XX Mycobacterium tuberculosis, and a heterologous polynucleotide sequence.
XX CC The recombinant fusion nucleic acids and polypeptides are useful for
XX providing stable and high yield expression of fusion polypeptides of both
XX eukaryotic and prokaryotic origin and to encode a protein product for use
XX as an antigen for detecting serum antibodies. The presence of serum

CC antibodies to M. tuberculosis antigens in an individual indicates that
CC the individual is infected with it. The fusion polypeptides are useful as
CC sources of proteins for monitoring binding of serum antibodies to fusion
CC proteins and as an immunogen to induce and/or enhance immune responses.
CC The coding sequences can be ligated with a coding sequence of another
CC molecule such as M. tuberculosis antigens, a cytokine or an adjuvant and
CC can be used in vivo as a DNA vaccine. This sequence represents the
CC Mycobacterium tuberculosis MTB32A protein
XX
SQ Sequence 355 AA;
Query Match 99.4%; Score 1792; DB 4; Length 355;
Best Local Similarity 99.4%; Pred. No. 4e-125;
Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNSRRSLRWLLSVLAAGVGLGATAPAAQAPALPALSQDRFADFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWLLSVLAAGVGLGATAPAAQAPALPALSQDRFADFPALPLDPSAMVAQV 60
QY 61 PQVNTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDVG 120
DB 61 PQVNTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDVG 120
QY 121 YRTQDVAVLQIRGAGLPSAAIGGVAVGEPVVMGNSGGGTPRAVGRVVALGQTV 180
DB 121 YRTQDVAVLQIRGAGLPSAAIGGVAVGEPVVMGNSGGGTPRAVGRVVALGQTV 180
QY 181 QASDSLTCGAETLNGLIQFDAAIQGDSGGPVVNGLGQVGMNTAASDNFOLSGGGGPA 240
DB 181 QASDSLTCGAETLNGLIQFDAAIQGDSGGPVVNGLGQVGMNTAASDNFOLSGGGGPA 240
QY 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVDNNGGARVQVVGSAPAASLGIST 300
DB 241 IPIQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVDNNGGARVQVVGSAPAASLGIST 300
QY 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAGGPPA 355
DB 301 GDVITAVDGAPINSATAMADALNGHHPGDVISVNWQTKSGGTRTGNVTLAGGPPA 355

RESULT 12
AAG81110
ID AAG81110 standard; protein; 355 AA.
XX AAG81110;
XX AC AAG81110;
XX AC AAG81110;
DT 04-SEP-2001 (first entry)
XX Mycobacterium tuberculosis potential drug target protein SEQ ID 161.
XX DE Mycobacterium tuberculosis potential drug target protein SEQ ID 161.
XX KW Drug target; growth; organism viability; characterisation.
XX OS Mycobacterium tuberculosis.
XX PN WO200135317-A1.
XX PD 17-MAY-2001.
XX PF 13-NOV-2000; 2000WO-US031152.
XX PR 12-NOV-1999; 99US-0165086P.
XX PR 12-NOV-1999; 99US-0165124P.
XX PR 01-FEB-2000; 2000US-0179531P.
XX PA (REGC) UNIV CALIFORNIA.
XX PI Eisenberg D, Rotstein SH, Marcotte EM;
XX WPI; 2001-329193/34.
XX DR N-PSDB; AAL40768.
XX PT Identifying nucleotide or polypeptide sequence for use as drug target,
XX involves providing algorithm that analyzes a functional relationship

PT between nucleotide or polypeptide sequences, and comparing the sequences.
 XX
 PS Disclosure; Page 157; 207pp; English.
 XX
 CC This invention relates to a method for identifying a nucleotide or
 CC polypeptide sequence that may be a drug target, or essential for growth
 CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
 CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
 CC tuberculosis proteins which are potential drug targets. The DNA and
 CC protein sequences are used to illustrate the method of the invention. The
 CC method involves providing an unknown nucleotide or polypeptide sequences,
 CC and comparing it to a number of sequences along with at least one
 CC algorithm capable of analysing a functional relationship between
 CC nucleotide and polypeptide sequences. The method is useful for
 CC characterising the function of nucleic acids and polypeptides that may be
 CC useful as a target for a drug or essential for the growth or viability of
 CC an organism
 XX
 SQ Sequence 355 AA;

Query Match 99.4%; Score 1792; DB 4; Length 355;
 Best Local Similarity 99.4%; Pred. No. 4e-125;
 Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MNSRRSLRWWSLLSVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60
 DB 1 MNSRRSLRWWSLLSVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQV 60
 QY 61 PQVNTINTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
 DB 61 PQVNTINTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
 QY 121 YDRTQDVAVQLRGAGGLPSAAIGGVAVGEPVWAMNSGGGGTTPRAVPGRVVALGQTV 180
 DB 121 YDRTQDVAVQLRGAGGLPSAAIGGVAVGEPVWAMNSGGGGTTPRAVPGRVVALGQTV 180
 QY 181 QASDSLTCABETLNGLIQFDAAIQPDGSGPVVNGLGQVWGMNTAASDNFOLSGGGQGFA 240
 DB 181 QASDSLTCABETLNGLIQFDAAIQPDGSGPVVNGLGQVWGMNTAASDNFOLSGGGQGFA 240
 QY 241 IPIGAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGARVQVVGSAAPASLGIST 300
 DB 241 IPIGAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGARVQVVGSAAPASLGIST 300
 QY 301 GDVITAVDGPINSATAMADALNGHPGDVIVSVNMQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGPINSATAMADALNGHPGDVIVSVNMQTKSGGTRTGNVTLAEGPPA 355

RESULT 13
 AA04830
 ID AA04830 standard; protein; 379 AA.
 XX
 AC AAY04830;
 XX
 DT 06-JUL-1999 (first entry)
 XX
 DE Mycobacterium species protein sequence 50F.
 XX
 KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 KW hybridisation; detection; vaccine; immunisation; infection.
 XX
 OS Mycobacterium sp.
 XX
 PN WO990186-A2.
 XX
 PD 25-FEB-1999.
 XX
 PF 14-AUG-1998; 98WO-FR001813.
 XX
 PR 14-AUG-1997; 97FR-00010404.
 PR 11-SEP-1997; 97FR-00011325.
 XX

PA (INSP) INST PASTEUR.
 XX
 PI Gicquel B, Portnoie D, Lim B, Pelicic V, Guigueno A;
 PI Goguet De La Salmoniere Y;
 XX
 DR WPI; 1999-181045/15.
 DR N-PSDB; AAX34252.
 XX
 PT Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in infection-associated
 PT protein expression.
 XX
 PS Claim 32; Fig 50F; 309pp; French.
 XX
 CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
 CC from various Mycobacterium species microorganisms. The encoding
 CC nucleotide sequences can be used as primers and probes for methods for
 CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection
 XX
 SQ Sequence 379 AA;

Query Match 99.4%; Score 1792; DB 2; Length 379;
 Best Local Similarity 99.4%; Pred. No. 4.3e-125;
 Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MNSRRSLRWWSLLSVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQVA 60
 DB 25 MNSRRSLRWWSLLSVLAAGVGLATAPAAQAAPPALSDQRFADFPALPLDPSAMVAQV 84
 QY 61 PQVNTINTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 120
 DB 85 PQVNTINTKLYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVVG 144
 QY 121 YDRTQDVAVQLRGAGGLPSAAIGGVAVGEPVWAMNSGGGGTTPRAVPGRVVALGQTV 180
 DB 145 YDRTQDVAVQLRGAGGLPSAAIGGVAVGEPVWAMNSGGGGTTPRAVPGRVVALGQTV 204
 QY 181 QASDSLTCABETLNGLIQFDAAIQPDGSGPVVNGLGQVWGMNTAASDNFOLSGGGQGFA 240
 DB 205 QASDSLTCABETLNGLIQFDAAIQPDGSGPVVNGLGQVWGMNTAASDNFOLSGGGQGFA 264
 QY 241 IPIGAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGARVQVVGSAAPASLGIST 300
 DB 265 IPIGAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGARVQVVGSAAPASLGIST 324
 QY 301 GDVITAVDGPINSATAMADALNGHPGDVIVSVNMQTKSGGTRTGNVTLAEGPPA 355
 DB 325 GDVITAVDGPINSATAMADALNGHPGDVIVSVNMQTKSGGTRTGNVTLAEGPPA 379

RESULT 14
 AA29702
 ID AA29702 standard; protein; 330 AA.
 XX
 AC AAE29702;
 XX
 DT 27-JAN-2003 (first entry)
 XX
 DE Mycobacterium tuberculosis mature Ra35 antigenic protein.
 XX
 KW Vaccine; immunity; diagnostic agent; gene therapy; Ra35 antigen.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PH Key Location/Qualifiers
 FT Misc-difference 182 /note= "Encoded by GAG"
 FT Misc-difference 183 /note= "Encoded by GCG"
 FT
 XX WO200272792-A2.

XX 19-SEP-2002.
 XX 13-MAR-2002; 2002WO-US008223.
 XX 13-MAR-2001; 2001US-0275837P.
 XX (CORI-) CORIXA CORP.
 XX Skeiky Y, Brannon M, Guderian J;
 XX WPI; 2002-759844/82.
 XX N-PSDB; AAD47077.
 XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF, M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity against pathogenic microorganisms e.g. Leishmania and Mycobacterium tuberculosis.
 XX Disclosure; Page 79-80; 155pp; English.
 XX The invention relates to a recombinant nucleic acid molecule encoding a fusion polypeptide. The recombinant nucleic acid comprises a heterologous polynucleotide sequence encoding an antigen or an antigenic fragment from Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a polypeptide or its fragment. The Leishmania polynucleotide is selected from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention are used in methods for eliciting immune response in mammals. They are useful as vaccines to elicit protective immunity against pathogenic microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion polypeptides are used for enhancing the expression of polynucleotides, as in vivo diagnostic agents and for raising antibodies in a non-human animal. The invention is used in gene therapy. The present sequence is M. tuberculosis mature Ra35 (N-terminus of MTB32A; RA35FL) antigenic protein
 XX Sequence 330 AA;
 SQ
 Query Match 90.7%; Score 1634; DB 5; Length 330;
 Best Local Similarity 99.4%; Pred. No. 2.1e-113;
 Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 33 APPALSQDRFADFPALPLDPSAMVAQVPOVNVNTKLGYNNAVAGTGIVIDNGVLT 92
 Db 8 APPALSQDRFADFPALPLDPSAMVAQVPOVNVNTKLGYNNAVAGTGIVIDNGVLT 67
 QY 93 NNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGLPSAAIIGGGVAVGEP 152
 Db 68 NNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGLPSAAIIGGGVAVGEP 127
 QY 153 VVAMGNSGGGTPRAVGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQDSDGGPV 212
 Db 128 VVAMGNSGGGTPRAVGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQDSDGGPV 187
 QY 213 VNLGQVVGNTAASDNFQSGGGQFAIPICQAMAIAGQIRSGGSPVTHIGTAFGL 272
 Db 188 VNLGQVVGNTAASDNFQSGGGQFAIPICQAMAIAGQIRSGGSPVTHIGTAFGL 247
 QY 273 GVVDNNGGARVQRVVGSAPASLIGISTGDVITAVDGPINSATAMADALNGHHPGDVIS 332
 Db 248 GVVDNNGGARVQRVVGSAPASLIGISTGDVITAVDGPINSATAMADALNGHHPGDVIS 307
 QY 333 VNWQTKSGTGTGNVTLAEGPPA 355
 Db 308 VTWQTKSGTGTGNVTLAEGPPA 330
 RESULT 15
 AAEL17566
 ID AAEL17566 standard; protein; 330 AA.
 XX
 AC AAEL17566;
 XX
 DT 22-APR-2002 (first entry)

XX Mycobacterium sp. MTB32A (Ra35FL) mature protein.
 DE Fusion protein; antigen; serological sensitivity; immune response;
 XX tuberculosis; infection; vaccine; MTB32A; Ra32FL protein.
 KW
 XX Mycobacterium sp.
 OS
 XX Key Location/Qualifiers
 FH Region 8..202
 FT /note= "Ra35 N-terminal peptide"
 FT Misc-difference 182
 FT /note= "Encoded by GAG"
 FT Misc-difference 183
 FT /note= "Encoded by GCG"
 FT Region 199..330
 FT /note= "Ra35 C-terminal peptide, Ra12"
 XX
 PN WO200198460-A2.
 XX
 XX 27-DEC-2001.
 XX
 XX 20-JUN-2001; 2001WO-US019959.
 PF
 XX 20-JUN-2001; 2000US-00597796.
 XX
 PR 01-FEB-2001; 2001US-0265737P.
 PR
 XX (CORI-) CORIXA CORP.
 PA
 XX Skeiky Y, Reed S, Alderson M;
 XX WPI; 2002-147798/19.
 FI
 XX N-PSDB; AAD28336.
 DR
 XX Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 XX Claim 70; Fig 6; 136pp; English.
 PS
 XX The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB32A (Ra32FL) mature protein
 CC
 XX Sequence 330 AA;
 SQ
 Query Match 90.7%; Score 1634; DB 5; Length 330;
 Best Local Similarity 99.4%; Pred. No. 2.1e-113;
 Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 33 APPALSQDRFADFPALPLDPSAMVAQVPOVNVNTKLGYNNAVAGTGIVIDNGVLT 92
 Db 8 APPALSQDRFADFPALPLDPSAMVAQVPOVNVNTKLGYNNAVAGTGIVIDNGVLT 67
 QY 93 NNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGLPSAAIIGGGVAVGEP 152
 Db 68 NNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGLPSAAIIGGGVAVGEP 127
 QY 153 VVAMGNSGGGTPRAVGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQDSDGGPV 212
 Db 128 VVAMGNSGGGTPRAVGRVVALGQTVQASDLSLTGAEETLNGLIQFDAAIQDSDGGPV 187
 QY 213 VNLGQVVGNTAASDNFQSGGGQFAIPICQAMAIAGQIRSGGSPVTHIGTAFGL 272
 Db 188 VNLGQVVGNTAASDNFQSGGGQFAIPICQAMAIAGQIRSGGSPVTHIGTAFGL 247
 QY 273 GVVDNNGGARVQRVVGSAPASLIGISTGDVITAVDGPINSATAMADALNGHHPGDVIS 332
 Db 248 GVVDNNGGARVQRVVGSAPASLIGISTGDVITAVDGPINSATAMADALNGHHPGDVIS 307
 QY 333 VNWQTKSGTGTGNVTLAEGPPA 355
 Db 308 VTWQTKSGTGTGNVTLAEGPPA 330

Db 68 NNHVIAGATDINAFSVSGGTYGYDVVGYDRTQDVAVIQLRGAGGLPSAAIGGVAVGEP 127
Qy 153 VVAMNGSGGGCTPRVPGVWVVALGQTVQASDSLTAETTLNGLIOFDAAIOPGDSGGPV 212
Db 128 VVAMNGSGGGCTPRVPGVWVVALGQTVQASDSLTAETTLNGLIOFDAAIOPGDSGGPV 187
Qy 213 VVGLQVVGWMTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGGSPVHIGPTAFLGL 272
Db 188 VVGLQVVGWMTAASDNFQLSQGGQGFAPIGQAMAIAGQIRSGGGSPVHIGPTAFLGL 247
Qy 273 GVVDNNGNGARVQVRVVGSAFASLIGISTGDIITAVDGAIPINSATAMADALNGHFGDVIS 332
Db 248 GVVDNNGNGARVQVRVVGSAFASLIGISTGDIITAVDGAIPINSATAMADALNGHFGDVIS 307
Qy 333 VVWQTKSGGTRTGNVTILAEPPA 355
Db 308 VVWQTKSGGTRTGNVTILAEPPA 330

Search completed: June 30, 2004, 16:48:46
Job time : 38.5039 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:44:57 ; Search time 10.7991 Seconds
(without alignments)
1697.113 Million cell updates/sec

Title: US-09-597-796C-4
Perfect score: 1802
Sequence: 1 MNSRRSLRWSLLSVLAA.....QTKSGTRGTGNTLAEGPPA 355

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/PCTUS COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1802	100.0	355	3	US-08-818-112-79
2	1802	100.0	355	4	US-08-818-111-80
3	1802	100.0	355	4	US-09-056-556-79
4	1802	100.0	355	4	US-09-072-596-80
5	1802	100.0	355	4	US-09-072-967-79
6	983	54.6	596	4	US-09-287-849-26
7	983	54.6	729	4	US-09-223-040-2
8	983	54.6	729	4	US-09-287-849-2
9	676	37.5	132	3	US-08-818-112-66
10	676	37.5	132	4	US-08-818-111-67
11	676	37.5	132	4	US-09-056-556-66
12	676	37.5	132	4	US-09-072-596-67
13	676	37.5	132	4	US-09-072-967-66
14	676	37.5	132	4	US-09-636-215-819
15	676	37.5	132	4	US-09-636-215-848
16	676	37.5	132	4	US-09-685-166A-819
17	676	37.5	132	4	US-09-685-166A-848
18	671	37.2	231	4	US-09-287-849-28
19	628	34.9	224	4	US-09-636-215-825
20	628	34.9	224	4	US-09-685-166A-825
21	628	34.9	267	4	US-09-643-597-352
22	628	34.9	267	4	US-09-606-421B-352
23	628	34.9	273	4	US-09-736-457-1864
24	628	34.9	299	4	US-09-643-597-354
25	628	34.9	299	4	US-09-606-421B-354
26	628	34.9	304	4	US-09-636-215-835
27	628	34.9	304	4	US-09-685-166A-835

Sequence 1863, App
Sequence 852, App
Sequence 852, App
Sequence 349, App
Sequence 349, App
Sequence 333, App
Sequence 333, App
Sequence 196, App
Sequence 196, App
Sequence 196, App
Sequence 353, App
Sequence 353, App
Sequence 337, App
Sequence 337, App
Sequence 309, App
Sequence 309, App
Sequence 325, App
Sequence 325, App

ALIGNMENTS

RESULT 1
US-08-818-112-79
; Sequence 79, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 79:
SEQUENCE CHARACTERISTICS:
LENGTH: 355 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-79

Query Match 100.0%; Score 1802; DB 3; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.1e-151;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNSRRSLRWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
 DB 1 MSNSRRSLRWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNTINTKLYNNVAGAGTGIVIDPVGVLVTTNNHVIAGATDINAFSVSGGQTYGVDDVVG 120
 DB 61 PQVNTINTKLYNNVAGAGTGIVIDPVGVLVTTNNHVIAGATDINAFSVSGGQTYGVDDVVG 120
 QY 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGGQGTTPRAVGRVVALGQTV 180
 DB 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGGQGTTPRAVGRVVALGQTV 180
 QY 181 QASDSLTTGAEETLNGLIQFDAAIQPGDSGPPVNGLVGVVGMNTAASDNFQLSQGGQGA 240
 DB 181 QASDSLTTGAEETLNGLIQFDAAIQPGDSGPPVNGLVGVVGMNTAASDNFQLSQGGQGA 240
 QY 241 IPTGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAFPAASLGIST 300
 DB 241 IPTGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAFPAASLGIST 300
 QY 301 GDVITAVDGPAPINSATAMADALNGHHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGPAPINSATAMADALNGHHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 2

US-08-818-111-80
 ; Sequence 80, Application US/08818111
 ; Patent No. 6338852
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonia
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 148
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED AND BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION NUMBER: US/08/818,111
 ; FILING DATE: 13-MAR-1997
 ; CLASSIFICATION: 424
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.417C6
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 80:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 355 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-08-818-111-80

Query Match 100.0%; Score 1802; DB 4; Length 355;
 Best Local Similarity 100.0%; Pred. No. 2.1e-151;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNSRRSLRWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
 DB 1 MSNSRRSLRWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60
 QY 61 PQVNTINTKLYNNVAGAGTGIVIDPVGVLVTTNNHVIAGATDINAFSVSGGQTYGVDDVVG 120
 DB 61 PQVNTINTKLYNNVAGAGTGIVIDPVGVLVTTNNHVIAGATDINAFSVSGGQTYGVDDVVG 120
 QY 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGGQGTTPRAVGRVVALGQTV 180
 DB 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGGQGTTPRAVGRVVALGQTV 180
 QY 181 QASDSLTTGAEETLNGLIQFDAAIQPGDSGPPVNGLVGVVGMNTAASDNFQLSQGGQGA 240
 DB 181 QASDSLTTGAEETLNGLIQFDAAIQPGDSGPPVNGLVGVVGMNTAASDNFQLSQGGQGA 240
 QY 241 IPTGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAFPAASLGIST 300
 DB 241 IPTGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGNGARVQVVGSAFPAASLGIST 300
 QY 301 GDVITAVDGPAPINSATAMADALNGHHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGPAPINSATAMADALNGHHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 3

US-09-056-556-79
 ; Sequence 79, Application US/09056556
 ; Patent No. 6350456
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
 ; NUMBER OF SEQUENCES: 241
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED AND BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION NUMBER: US/09/056,556
 ; FILING DATE: 07-APR-1998
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.457
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 79:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 355 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-09-056-556-79

Query Match 100.0%; Score 1802; DB 4; Length 355;
 Best Local Similarity 100.0%; Pred. No. 2.1e-151;
 Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNSRRSLRWSLLSVLAAGVGLATAPAAQAPALSDQRFADFPALPLDPSAMVAQVA 60

Db 1 MNSRRSLRWSLLVLAAYGLGATAPAAQAPALSDRFPADFPALPLDPSAMVAQVA 60
QY 61 PQVWINTKLGYNNAVGAGTGVIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVDVG 120
Db 61 PQVWINTKLGYNNAVGAGTGVIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVDVG 120
QY 121 YDRTQDVAVLQRLGAGGLPSAAIGGGVAVGPPVAMGNSGGGTPRAVPGRVVALGQTV 180
Db 121 YDRTQDVAVLQRLGAGGLPSAAIGGGVAVGPPVAMGNSGGGTPRAVPGRVVALGQTV 180
QY 181 QASDSLTAETLNGLIQFDDAIQPDGSGPVVNGLGQVGMNTAASDNFOLSGGGGFA 240
Db 181 QASDSLTAETLNGLIQFDDAIQPDGSGPVVNGLGQVGMNTAASDNFOLSGGGGFA 240
QY 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVQVVGSAPAAASLGIST 300
Db 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVQVVGSAPAAASLGIST 300
QY 301 GDVITAVDGAIPNSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
Db 301 GDVITAVDGAIPNSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 4

US-09-072-596-80
; Sequence 80, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSES: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCES/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 80:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-80

Query Match 100.0%; Score 1802; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.1e-151;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWSLLVLAAYGLGATAPAAQAPALSDRFPADFPALPLDPSAMVAQVA 60
Db 1 MNSRRSLRWSLLVLAAYGLGATAPAAQAPALSDRFPADFPALPLDPSAMVAQVA 60
QY 61 PQVWINTKLGYNNAVGAGTGVIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVDVG 120
Db 61 PQVWINTKLGYNNAVGAGTGVIVDPNGVLTNNHVIAGATDINAFVSGSGQTYGVVDVG 120
QY 121 YDRTQDVAVLQRLGAGGLPSAAIGGGVAVGPPVAMGNSGGGTPRAVPGRVVALGQTV 180
Db 121 YDRTQDVAVLQRLGAGGLPSAAIGGGVAVGPPVAMGNSGGGTPRAVPGRVVALGQTV 180
QY 181 QASDSLTAETLNGLIQFDDAIQPDGSGPVVNGLGQVGMNTAASDNFOLSGGGGFA 240
Db 181 QASDSLTAETLNGLIQFDDAIQPDGSGPVVNGLGQVGMNTAASDNFOLSGGGGFA 240
QY 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVQVVGSAPAAASLGIST 300
Db 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFLGLGVVDNNGNGARVQVVGSAPAAASLGIST 300
QY 301 GDVITAVDGAIPNSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
Db 301 GDVITAVDGAIPNSATAMADALNGHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 5

US-09-072-967-79
; Sequence 79, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSES: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCES/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 79:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 355 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-967-79

```
Query Match 100.0%; Score 1802; DB 4; Length 355;
Best Local Similarity 100.0%; Pred. No. 2.1e-151; Indels 0; Gaps 0;
Matches 355; Conservative 0; Mismatches 0;

QY 1 MNSRRSLRWSLLSVAAGVGLGATAPAAQAPALSDRFPALPLDPSAMVAQVA 60
Db 1 MNSRRSLRWSLLSVAAGVGLGATAPAAQAPALSDRFPALPLDPSAMVAQVA 60

QY 61 PQVNTNKLGNNAAGCTGIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDVVG 120
Db 61 PQVNTNKLGNNAAGCTGIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDVVG 120

QY 121 YDRTQDVAVLQLRGAGLPSAIGGVAVGEPPVAMNSGGGGTTPRAVPGRWALGTV 180
Db 121 YDRTQDVAVLQLRGAGLPSAIGGVAVGEPPVAMNSGGGGTTPRAVPGRWALGTV 180

QY 181 QASDSLTGAETLNGLIQDAIQDSCGPPVNGLGQVVGMMNTAASDNFQSGGGQFPA 240
Db 181 QASDSLTGAETLNGLIQDAIQDSCGPPVNGLGQVVGMMNTAASDNFQSGGGQFPA 240

QY 241 IPIQAMALAGQIRSGGSGPTVHIGTAFGLGVVDNNGGARVORVVGSAAPASLGIST 300
Db 241 IPIQAMALAGQIRSGGSGPTVHIGTAFGLGVVDNNGGARVORVVGSAAPASLGIST 300

QY 301 GDVITAVDGAPINSATAMADALNGHHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
Db 301 GDVITAVDGAPINSATAMADALNGHHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 6
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 54.6%; Score 983; DB 4; Length 596;
Best Local Similarity 97.5%; Pred. No. 1.1e-78;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 APAQAAPPALSDRFPALPLDPSAMVAQVAPQVNTNKLGNNAAGCTGIVDPN 87
Db 397 AAGDIAPALSDRFPALPLDPSAMVAQVAPQVNTNKLGNNAAGCTGIVDPN 456

RESULT 8
US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US

QY 88 GVLTNNHVIAGATDINAFSGSGQTYGVDVVGVDRTQDVAVLQLRGAGLPSAAIGGV 147
Db 457 GVLTNNHVIAGATDINAFSGSGQTYGVDVVGVDRTQDVAVLQLRGAGLPSAAIGGV 516

QY 148 AVGEPVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDDAIQPGD 207
Db 517 AVGEPVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDDAIQPGD 576

QY 208 SGGPVVNGLGQVVGMMNTAAS 227
Db 577 SGGPVVNGLGQVVGMMNTAAS 596

RESULT 7
US-09-223-040-2
; Sequence 2, Application US/09223040
; Patent No. 6544522
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-223-040-2

Query Match 54.6%; Score 983; DB 4; Length 729;
Best Local Similarity 97.5%; Pred. No. 1.5e-78;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 APAQAAPPALSDRFPALPLDPSAMVAQVAPQVNTNKLGNNAAGCTGIVDPN 87
Db 530 AAGDIAPALSDRFPALPLDPSAMVAQVAPQVNTNKLGNNAAGCTGIVDPN 589

QY 88 GVLTNNHVIAGATDINAFSGSGQTYGVDVVGVDRTQDVAVLQLRGAGLPSAAIGGV 147
Db 590 GVLTNNHVIAGATDINAFSGSGQTYGVDVVGVDRTQDVAVLQLRGAGLPSAAIGGV 649

QY 148 AVGEPVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDDAIQPGD 207
Db 650 AVGEPVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDDAIQPGD 709

QY 208 SGGPVVNGLGQVVGMMNTAAS 227
Db 710 SGGPVVNGLGQVVGMMNTAAS 729

RESULT 8
US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
```

CURRENT APPLICATION NUMBER: US/09/287,849
CURRENT FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 729
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-2

Query Match 54.6%; Score 983; DB 4; Length 729;
Best Local Similarity 97.5%; Pred. No. 1.5e-78;
Matches 195; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 28 APAQAAPALSDRFADFPALPLDPSAWAQVAPQVNVNINIKLGYNNVAGAGTGIVIDPN 87
DB 530 AAGDIAPALSDRFADFPALPLDPSAWAQVAPQVNVNINIKLGYNNVAGAGTGIVIDPN 589
QY 88 GVLITNNHVIAGATDINAFSGSGQTYGVDVVGVDRTQDVAVIQLRGAGGLPSAAIGGV 147
DB 590 GVLITNNHVIAGATDINAFSGSGQTYGVDVVGVDRTQDVAVIQLRGAGGLPSAAIGGV 649
QY 148 AVGEFVAMGNSGCGGTTPRAVGRVVALGQTVQASDLSLTCAEETLNGLIQFRAAIQPGD 207
DB 650 AVGEFVAMGNSGCGGTTPRAVGRVVALGQTVQASDLSLTCAEETLNGLIQFRAAIQPGD 709
QY 208 SGGFVWNGLGQVVGWNTAAS 227
DB 710 SGGFVWNGLGQVVGWNTAAS 729

RESULT 9
US-08-818-112-66
Sequence 66, Application US/08818112
Patent No. 6290369
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J. 31,392
REGISTRATION NUMBER: 210121.411C6
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-66

Query Match 37.5%; Score 676; DB 3; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASNFOLSQGGQGAIPIGQAMAIAQIRSGGSPVHIGPTAFLGLGVVDNNGGAR 283
DB 1 TAASNFOLSQGGQGAIPIGQAMAIAQIRSGGSPVHIGPTAFLGLGVVDNNGGAR 60
QY 284 VORVVGSPAASLGISTGDTVITAVDGPINSTATADALNGHHPGDVIVSNWQTSGGTR 343
DB 61 VORVVGSPAASLGISTGDTVITAVDGPINSTATADALNGHHPGDVIVSNWQTSGGTR 120
QY 344 TGNVTLAEGPPA 355
DB 121 TGNVTLAEGPPA 132

RESULT 10
US-08-818-111-67
Sequence 67, Application US/08818111
Patent No. 6338852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J. 31,392
REGISTRATION NUMBER: 210121.417C6
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 67:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids

TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-67

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASDNFOLSQGGQGAFAIPGQAMAIAQIRSGGSPVTHIGTAFGLGVVDNNGGAR 283
DB 1 TAASDNFOLSQGGQGAFAIPGQAMAIAQIRSGGSPVTHIGTAFGLGVVDNNGGAR 60

QY 284 VORVVGSAAPASLGISTGDTAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTR 343
DB 61 VORVVGSAAPASLGISTGDTAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTR 120

QY 344 TGNVTLAEGPPA 355
DB 121 TGNVTLAEGPPA 132

RESULT 11
US-09-056-556-66
; Sequence 66, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSER: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-056-556-66

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASDNFOLSQGGQGAFAIPGQAMAIAQIRSGGSPVTHIGTAFGLGVVDNNGGAR 283
DB 1 TAASDNFOLSQGGQGAFAIPGQAMAIAQIRSGGSPVTHIGTAFGLGVVDNNGGAR 60

QY 284 VORVVGSAAPASLGISTGDTAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTR 343
DB 61 VORVVGSAAPASLGISTGDTAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTR 120

DB 61 VORVVGSAAPASLGISTGDTAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTR 120

QY 344 TGNVTLAEGPPA 355
DB 121 TGNVTLAEGPPA 132

RESULT 12
US-09-072-596-67
; Sequence 67, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSER: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 67:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-596-67

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 224 TAASDNFOLSQGGQGAFAIPGQAMAIAQIRSGGSPVTHIGTAFGLGVVDNNGGAR 283
DB 1 TAASDNFOLSQGGQGAFAIPGQAMAIAQIRSGGSPVTHIGTAFGLGVVDNNGGAR 60

QY 284 VORVVGSAAPASLGISTGDTAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTR 343
DB 61 VORVVGSAAPASLGISTGDTAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTR 120

QY 344 TGNVTLAEGPPA 355
DB 121 TGNVTLAEGPPA 132

RESULT 13
US-09-072-967-66

Sequence 66, Application US/09072967
Patent No. 6592877
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Twardzik, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/072,967
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 66:
SEQUENCE CHARACTERISTICS:
LENGTH: 132 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-967-66
Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 224 TAASDNFOLSGGGGFAIPGQMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGGAR 283
Db 1 TAASDNFOLSGGGGFAIPGQMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGGAR 60
QY 284 VORVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTR 343
Db 61 VORVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTR 120
QY 344 TGNVTLAEGPPA 355
Db 121 TGNVTLAEGPPA 132
RESULT 14
US-09-636-215-819
Sequence 819, Application US/09636215
Patent No. 6620922
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqui
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.42717C17
CURRENT APPLICATION NUMBER: US/09/636,215
CURRENT FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 852
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 848
LENGTH: 132

APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.42717C17
CURRENT APPLICATION NUMBER: US/09/636,215
CURRENT FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 852
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 819
LENGTH: 132
TYPE: PRT
ORGANISM: Homo sapien
US-09-636-215-819
Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 224 TAASDNFOLSGGGGFAIPGQMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGGAR 283
Db 1 TAASDNFOLSGGGGFAIPGQMAIAGQIRSGGSPVHIGPTAFGLGVVDNNGGAR 60
QY 284 VORVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTR 343
Db 61 VORVVGSAAPASLGISTGDTVITAVDGPINSATAMADALNGHHPGDVISVNWQTKSGGTR 120
QY 344 TGNVTLAEGPPA 355
Db 121 TGNVTLAEGPPA 132
RESULT 15
US-09-636-215-848
Sequence 848, Application US/09636215
Patent No. 6620922
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan L.
APPLICANT: Jiang, Yuqui
APPLICANT: Henderson, Robert A.
APPLICANT: Kalos, Michael D.
APPLICANT: Fanger, Gary R.
APPLICANT: Retter, Marc W.
APPLICANT: Stolk, John A.
APPLICANT: Day, Craig H.
APPLICANT: Vedvick, Thomas S.
APPLICANT: Carter, Darrick
APPLICANT: Li, Samuel
APPLICANT: Wang, Aijun
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Hepler, William
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.42717C17
CURRENT APPLICATION NUMBER: US/09/636,215
CURRENT FILING DATE: 2000-08-10
NUMBER OF SEQ ID NOS: 852
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 848
LENGTH: 132

TYPE: PRT
ORGANISM: Homo sapiens
US-09-636-215-848

Query Match 37.5%; Score 676; DB 4; Length 132;
Best Local Similarity 100.0%; Pred. No. 2e-52;
Matches 132; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	224	TAASDNFQLSQGGGPAIPIGQAMAIAGQIRSGGGSPVHIGPTAFLGLGVVDNNGGAR	283
Db	1	TAASDNFQLSQGGGPAIPIGQAMAIAGQIRSGGGSPVHIGPTAFLGLGVVDNNGGAR	60
QY	284	VQRVVGSAAPASLGISTGDIVITAVDGAPINSATAMADALNGHHPGDVISVNMQTKSGGTR	343
Db	61	VQRVVGSAAPASLGISTGDIVITAVDGAPINSATAMADALNGHHPGDVISVNMQTKSGGTR	120
QY	344	TGNVTLAEGPPA	355
Db	121	TGNVTLAEGPPA	132

Search completed: June 30, 2004, 16:55:18
Job time : 11.7991 secs

QY 121 YDRTQDVAVLQRLGAGLPSAATGGGVAVGEPVVMNGSGGGGTTPRAVGRVVALGQTV 180
DB 121 YDRTQDVAVLQRLGAGLPSAATGGGVAVGEPVVMNGSGGGGTTPRAVGRVVALGQTV 180
QY 181 QASDSLTAETLNGLIQFDAAIQPDGSGPVVNGLGQVVGWNTAASDNFOLSGGGQGA 240
DB 181 QASDSLTAETLNGLIQFDAAIQPDGSGPVVNGLGQVVGWNTAASDNFOLSGGGQGA 240
QY 241 IPIGQAMAIAGQIRSGGSPTHVIGTAFGLGVVDNNGNGARVQVVGSAAPASLGIST 300
DB 241 IPIGQAMAIAGQIRSGGSPTHVIGTAFGLGVVDNNGNGARVQVVGSAAPASLGIST 300
QY 301 GDVITAVDGPINSATAMADALNGHHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGPINSATAMADALNGHHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 2

US-10-193-002-80

; Sequence 80, Application US/10193002

; Publication No. US20030135026A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; Skeiky, Yasir A.W.

; Dillon, Davin C.

; Campos-Neto, Antonia

; Houghton, Raymond

; Vedvick, Thomas S.

; Twardzik, Daniel R.

; Lodes, Michael J.

; Hendrickson, Ronald C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

; TUBERCULOSIS

; NUMBER OF SEQUENCES: 350

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/193,002

; FILING DATE: 10-Jul-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/072,596

; FILING DATE: 05-MAY-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.417C9

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 80:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 355 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; SEQUENCE DESCRIPTION: SEQ ID NO: 80:

US-10-193-002-80

Query Match

Best Local Similarity 100.0%; Score 1802; DB 14; Length 355;

Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSNSRRSLRWSMLLSVLAAGVGLGATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60
DB 1 MSNSRRSLRWSMLLSVLAAGVGLGATAPAAQAAPPALSQDRFADFPALPLDPSAMVAQVA 60
QY 61 PQVWNINKLYNNNAVAGAGTIVIDPENGVLVTNNHVIAGATDINAFSVSGGQYGVDDVVG 120
DB 61 PQVWNINKLYNNNAVAGAGTIVIDPENGVLVTNNHVIAGATDINAFSVSGGQYGVDDVVG 120
QY 121 YDRTQDVAVLQRLGAGLPSAATGGGVAVGEPVVMNGSGGGGTTPRAVGRVVALGQTV 180
DB 121 YDRTQDVAVLQRLGAGLPSAATGGGVAVGEPVVMNGSGGGGTTPRAVGRVVALGQTV 180
QY 181 QASDSLTAETLNGLIQFDAAIQPDGSGPVVNGLGQVVGWNTAASDNFOLSGGGQGA 240
DB 181 QASDSLTAETLNGLIQFDAAIQPDGSGPVVNGLGQVVGWNTAASDNFOLSGGGQGA 240
QY 241 IPIGQAMAIAGQIRSGGSPTHVIGTAFGLGVVDNNGNGARVQVVGSAAPASLGIST 300
DB 241 IPIGQAMAIAGQIRSGGSPTHVIGTAFGLGVVDNNGNGARVQVVGSAAPASLGIST 300
QY 301 GDVITAVDGPINSATAMADALNGHHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGPINSATAMADALNGHHHPGDVIVSNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 3

US-10-084-843-79

; Sequence 79, Application US/10084843

; Publication No. US20030143243A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; Skeiky, Yasir A.W.

; Dillon, Davin C.

; Campos-Neto, Antonio

; Houghton, Raymond

; Vedvick, Thomas S.

; Twardzik, Daniel R.

; Lodes, Michael J.

; Hendrickson, Ronald C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

; AND DIAGNOSIS OF TUBERCULOSIS

; NUMBER OF SEQUENCES: 355

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED and BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/084,843

; FILING DATE: 25-Feb-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/09/072,967

; FILING DATE: 05-MAY-1998

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.411C9

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 79:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 355 amino acids

; TYPE: amino acid


```

; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 79;
US-10-084-843-79

Query Match      100.0%; Score 1802; DB 14; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.6e-134;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWWSLLSVAAGVGLGATAPAAQAPALSSQDRFADFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWWSLLSVAAGVGLGATAPAAQAPALSSQDRFADFPALPLDPSAMVAQVA 60
QY 61 PQVWNTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDVVG 120
DB 61 PQVWNTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDVVG 120
QY 121 YDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMNGSGGQGTTPRAVGRVVALGQTV 180
DB 121 YDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMNGSGGQGTTPRAVGRVVALGQTV 180
QY 181 QASDSLTCGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMTAAASDNFQLSGGGQGFA 240
DB 181 QASDSLTCGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMTAAASDNFQLSGGGQGFA 240
QY 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGARVQVRVVGSAAPASLGIST 300
DB 241 IPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVDNNGGARVQVRVVGSAAPASLGIST 300
QY 301 GDVITAVDGAPINSATAMADALNGHHGPDVLSVNWQTKSGGTRTGNVTLAEGPPA 355
DB 301 GDVITAVDGAPINSATAMADALNGHHGPDVLSVNWQTKSGGTRTGNVTLAEGPPA 355

RESULT 4
US-10-098-732A-2
; Sequence 2, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: 014058-C120100S
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; TYPE: PRT
; LENGTH: 355
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB32A (Ra35FL)
US-10-098-732A-2

Query Match      100.0%; Score 1802; DB 14; Length 355;
Best Local Similarity 100.0%; Pred. No. 1.6e-134;
Matches 355; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNSRRSLRWWSLLSVAAGVGLGATAPAAQAPALSSQDRFADFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWWSLLSVAAGVGLGATAPAAQAPALSSQDRFADFPALPLDPSAMVAQVA 60
QY 61 PQVWNTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDVVG 120
DB 61 PQVWNTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDVVG 120
QY 121 YDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMNGSGGQGTTPRAVGRVVALGQTV 180
DB 121 YDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMNGSGGQGTTPRAVGRVVALGQTV 180

RESULT 5
US-09-712-363-161
; Sequence 161, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 161
; LENGTH: 355
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: MTB363-161
US-09-712-363-161

Query Match      99.4%; Score 1792; DB 9; Length 355;
Best Local Similarity 99.4%; Pred. No. 9.6e-134;
Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNSRRSLRWWSLLSVAAGVGLGATAPAAQAPALSSQDRFADFPALPLDPSAMVAQVA 60
DB 1 MNSRRSLRWWSLLSVAAGVGLGATAPAAQAPALSSQDRFADFPALPLDPSAMVAQVA 60
QY 61 PQVWNTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDVVG 120
DB 61 PQVWNTKLGYNNAVAGTGVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDVVG 120
QY 121 YDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMNGSGGQGTTPRAVGRVVALGQTV 180
DB 121 YDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPVVMNGSGGQGTTPRAVGRVVALGQTV 180
QY 181 QASDSLTCGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMTAAASDNFQLSGGGQGFA 240
DB 181 QASDSLTCGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMTAAASDNFQLSGGGQGFA 240

```

Db 181 QASDLTGABETLNGLIQFDAAIQDGSQGVVNGLVGVVGMNTAASDNFQLSQGGGPA 240
QY 241 IPTQAMAIAGQIRSGGSGTTHIGTAFGLGVVDNNGNGARVQVWVGSAPAAISLGIST 300
Db 241 IPTQAMAIAGQIRSGGSGTTHIGTAFGLGVVDNNGNGARVQVWVGSAPAAISLGIST 300
QY 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWOTKSGGTRTGNVTILAEGPPA 355
Db 301 GDVITAVDGPINSATAMADALNGHHPGDVISVNWOTKSGGTRTGNVTILAEGPPA 355

RESULT 6
US-09-886-349A-4
; Sequence 4, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-0090703
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB32A (Ra35 mature)

US-09-886-349A-4

Query Match 90.7%; Score 1634; DB 12; Length 330;
Best Local Similarity 99.4%; Pred. No. 2.9e-121;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNVNINIKLGYNNVAGAGTGIVIDPNGVLT 92
Db 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNVNINIKLGYNNVAGAGTGIVIDPNGVLT 67
QY 93 NNHVIAGATDINAFSVGSGQTYGVVDVVDVVDVVDVVDVVDVVDVVDVVDVVDVVDV 152
Db 68 NNHVIAGATDINAFSVGSGQTYGVVDVVDVVDVVDVVDVVDVVDVVDVVDVVDVVDV 127
QY 153 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQDGSQGPV 212
Db 128 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQDGSQGPV 187
QY 213 VNLGVGVGMNTAASDNFQLSQGGGFAIPICQAMAIAGQIRSGGSGTTHIGTAFGL 272
Db 188 VNLGVGVGMNTAASDNFQLSQGGGFAIPICQAMAIAGQIRSGGSGTTHIGTAFGL 247
QY 273 GVVDNNGNGARVQVWVGSAPAAISLGISTGDTVITAVDGPINSATAMADALNGHHPGDVIS 332
Db 248 GVVDNNGNGARVQVWVGSAPAAISLGISTGDTVITAVDGPINSATAMADALNGHHPGDVIS 307
QY 333 VNWOTKSGGTRTGNVTILAEGPPA 355
Db 308 VTWQTKSGGTRTGNVTILAEGPPA 330

RESULT 7
US-10-098-732A-4
; Sequence 4, Application US/10098732A
; Publication No. US2003017529A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir

; APPLICANT: Brannon, Mark
; APPLICANT: Guderman, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Ra35 mature

US-10-098-732A-4

Query Match 90.7%; Score 1634; DB 14; Length 330;
Best Local Similarity 99.4%; Pred. No. 2.9e-121;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNVNINIKLGYNNVAGAGTGIVIDPNGVLT 92
Db 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNVNINIKLGYNNVAGAGTGIVIDPNGVLT 67
QY 93 NNHVIAGATDINAFSVGSGQTYGVVDVVDVVDVVDVVDVVDVVDVVDVVDVVDVVDV 152
Db 68 NNHVIAGATDINAFSVGSGQTYGVVDVVDVVDVVDVVDVVDVVDVVDVVDVVDVVDV 127
QY 153 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQDGSQGPV 212
Db 128 VVAMNSGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQDGSQGPV 187
QY 213 VNLGVGVGMNTAASDNFQLSQGGGFAIPICQAMAIAGQIRSGGSGTTHIGTAFGL 272
Db 188 VNLGVGVGMNTAASDNFQLSQGGGFAIPICQAMAIAGQIRSGGSGTTHIGTAFGL 247
QY 273 GVVDNNGNGARVQVWVGSAPAAISLGISTGDTVITAVDGPINSATAMADALNGHHPGDVIS 332
Db 248 GVVDNNGNGARVQVWVGSAPAAISLGISTGDTVITAVDGPINSATAMADALNGHHPGDVIS 307
QY 333 VNWOTKSGGTRTGNVTILAEGPPA 355
Db 308 VTWQTKSGGTRTGNVTILAEGPPA 330

RESULT 8
US-10-369-983-19
; Sequence 19, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderman, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: wild-type mature MTB32A (Ra35)

US-10-369-983-19

Query Match 90.7%; Score 1634; DB 15; Length 330;
Best Local Similarity 99.4%; Pred. No. 2.9e-121;
Matches 321; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLYNNAYCAGTGVVIDPENGVLIT 92
DB 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLYNNAYCAGTGVVIDPENGVLIT 67
QY 93 NNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
DB 68 NNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
QY 153 VVAMGNSGGGCTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
DB 128 VVAMGNSGGGCTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
QY 213 VNGLGQVVMNTAASDNFQSLQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGTAFGL 272
DB 188 VNGLGQVVMNTAASDNFQSLQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGTAFGL 247
QY 273 GVVDDNNGGARQVRVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHPGDIS 332
DB 248 GVVDDNNGGARQVRVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHPGDIS 307
QY 333 VNWQTKSGGTRTGNVTTLAEGPPA 355
DB 308 VTWQTKSGGTRTGNVTTLAEGPPA 330

RESULT 9

US-09-886-349A-6

; Sequence 6, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCES: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Ra35FLMutSA

Query Match 90.5%; Score 1631; DB 12; Length 330;
Best Local Similarity 99.1%; Pred. No. 4.9e-121;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLYNNAYCAGTGVVIDPENGVLIT 92
DB 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLYNNAYCAGTGVVIDPENGVLIT 67
QY 93 NNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
DB 68 NNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
QY 153 VVAMGNSGGGCTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
DB 128 VVAMGNSGGGCTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187

QY 213 VNGLGQVVMNTAASDNFQSLQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGTAFGL 272
DB 188 VNGLGQVVMNTAASDNFQSLQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGTAFGL 247
QY 273 GVVDDNNGGARQVRVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHPGDIS 332
DB 248 GVVDDNNGGARQVRVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHPGDIS 307
QY 333 VNWQTKSGGTRTGNVTTLAEGPPA 355
DB 308 VTWQTKSGGTRTGNVTTLAEGPPA 330

RESULT 10

US-10-098-732A-6

; Sequence 6, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 6
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Ra35FLMutSA

Query Match 90.5%; Score 1631; DB 14; Length 330;
Best Local Similarity 99.1%; Pred. No. 4.9e-121;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLYNNAYCAGTGVVIDPENGVLIT 92
DB 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLYNNAYCAGTGVVIDPENGVLIT 67
QY 93 NNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 152
DB 68 NNHVIAGATDINAFSVGSGQTYGVVDVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEP 127
QY 153 VVAMGNSGGGCTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
DB 128 VVAMGNSGGGCTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
QY 213 VNGLGQVVMNTAASDNFQSLQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGTAFGL 272
DB 188 VNGLGQVVMNTAASDNFQSLQGGGFAIPIGQAMAIAGQIRSGGSPTVHIGTAFGL 247
QY 273 GVVDDNNGGARQVRVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHPGDIS 332
DB 248 GVVDDNNGGARQVRVVGSAAPASLGISTGDIVITAVDGPINSATAMADALNGHHPGDIS 307
QY 333 VNWQTKSGGTRTGNVTTLAEGPPA 355
DB 308 VTWQTKSGGTRTGNVTTLAEGPPA 330

RESULT 11

US-10-369-983-20

; Sequence 20, Application US/10369983
; Publication No. US20030235593A1

```
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR FILING DATE: 2003-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mutated
; OTHER INFORMATION: MTB32MutSA (Ra35 mutSA)
US-10-369-983-20

Query Match          90.5%; Score 1631; DB 15; Length 330;
Best Local Similarity 99.1%; Pred. No. 4.9e-121;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVIVDPNGVVL 92
Db 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVIVDPNGVVL 67
Qy 93 NNHVIAGATDINAFSVSGSGTGYVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 152
Db 68 NNHVIAGATDINAFSVSGSGTGYVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 127
Qy 153 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
Db 128 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
Qy 213 VNLGQVVGWNTAAADNFQLSGGGQGFALPIGQMAIAGQIRSGGSPVTHIGPTAFI 272
Db 188 VNLGQVVGWNTAAADNFQLSGGGQGFALPIGQMAIAGQIRSGGSPVTHIGPTAFI 247
Qy 273 GVVDNNGARVORVVGSAAPASLIGTGDVITAVDGAIPINSATAMADALNGHHHPGDVIS 332
Db 248 GVVDNNGARVORVVGSAAPASLIGTGDVITAVDGAIPINSATAMADALNGHHHPGDVIS 307
Qy 333 VNWQTSKGGTRTGNVTLAEGPPA 355
Db 308 VTWQTSKGGTRTGNVTLAEGPPA 330

RESULT 12
US-10-369-983-2
; Sequence 2, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR FILING DATE: 2003-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 723
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mutated
; OTHER INFORMATION: MTB32MutSA (Ra35 mutSA)
US-10-369-983-20

Query Match          90.5%; Score 1631; DB 15; Length 330;
Best Local Similarity 99.1%; Pred. No. 4.9e-121;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVIVDPNGVVL 92
Db 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVIVDPNGVVL 67
Qy 93 NNHVIAGATDINAFSVSGSGTGYVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 152
Db 68 NNHVIAGATDINAFSVSGSGTGYVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 127
Qy 153 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
Db 128 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
Qy 213 VNLGQVVGWNTAAADNFQLSGGGQGFALPIGQMAIAGQIRSGGSPVTHIGPTAFI 272
Db 188 VNLGQVVGWNTAAADNFQLSGGGQGFALPIGQMAIAGQIRSGGSPVTHIGPTAFI 247
Qy 273 GVVDNNGARVORVVGSAAPASLIGTGDVITAVDGAIPINSATAMADALNGHHHPGDVIS 332
Db 248 GVVDNNGARVORVVGSAAPASLIGTGDVITAVDGAIPINSATAMADALNGHHHPGDVIS 307
Qy 333 VNWQTSKGGTRTGNVTLAEGPPA 355
Db 308 VTWQTSKGGTRTGNVTLAEGPPA 330

RESULT 12
US-10-369-983-2
; Sequence 2, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR FILING DATE: 2003-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 723
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mutated
; OTHER INFORMATION: MTB32MutSA (Ra35 mutSA)
US-10-369-983-20

Query Match          90.5%; Score 1631; DB 15; Length 330;
Best Local Similarity 99.1%; Pred. No. 4.9e-121;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVIVDPNGVVL 92
Db 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVIVDPNGVVL 67
Qy 93 NNHVIAGATDINAFSVSGSGTGYVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 152
Db 68 NNHVIAGATDINAFSVSGSGTGYVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 127
Qy 153 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
Db 128 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
Qy 213 VNLGQVVGWNTAAADNFQLSGGGQGFALPIGQMAIAGQIRSGGSPVTHIGPTAFI 272
Db 188 VNLGQVVGWNTAAADNFQLSGGGQGFALPIGQMAIAGQIRSGGSPVTHIGPTAFI 247
Qy 273 GVVDNNGARVORVVGSAAPASLIGTGDVITAVDGAIPINSATAMADALNGHHHPGDVIS 332
Db 248 GVVDNNGARVORVVGSAAPASLIGTGDVITAVDGAIPINSATAMADALNGHHHPGDVIS 307
Qy 333 VNWQTSKGGTRTGNVTLAEGPPA 355
Db 308 VTWQTSKGGTRTGNVTLAEGPPA 330
```

```
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR FILING DATE: 2003-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mutated
; OTHER INFORMATION: MTB32MutSA (Ra35 mutSA)
US-10-369-983-20

Query Match          90.5%; Score 1631; DB 15; Length 330;
Best Local Similarity 99.1%; Pred. No. 1.3e-120;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVIVDPNGVVL 92
Db 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVIVDPNGVVL 67
Qy 93 NNHVIAGATDINAFSVSGSGTGYVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 152
Db 68 NNHVIAGATDINAFSVSGSGTGYVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 127
Qy 153 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
Db 128 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
Qy 213 VNLGQVVGWNTAAADNFQLSGGGQGFALPIGQMAIAGQIRSGGSPVTHIGPTAFI 272
Db 188 VNLGQVVGWNTAAADNFQLSGGGQGFALPIGQMAIAGQIRSGGSPVTHIGPTAFI 247
Qy 273 GVVDNNGARVORVVGSAAPASLIGTGDVITAVDGAIPINSATAMADALNGHHHPGDVIS 332
Db 248 GVVDNNGARVORVVGSAAPASLIGTGDVITAVDGAIPINSATAMADALNGHHHPGDVIS 307
Qy 333 VNWQTSKGGTRTGNVTLAEGPPA 355
Db 308 VTWQTSKGGTRTGNVTLAEGPPA 330

RESULT 13
US-10-369-983-4
; Sequence 4, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR FILING DATE: 2003-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB-102F fusion
; OTHER INFORMATION: protein
US-10-369-983-4

Query Match          90.5%; Score 1631; DB 15; Length 1010;
Best Local Similarity 99.1%; Pred. No. 2e-120;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVIVDPNGVVL 92
Db 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVIVDPNGVVL 67
Qy 93 NNHVIAGATDINAFSVSGSGTGYVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 152
Db 68 NNHVIAGATDINAFSVSGSGTGYVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 127
Qy 153 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
Db 128 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
Qy 213 VNLGQVVGWNTAAADNFQLSGGGQGFALPIGQMAIAGQIRSGGSPVTHIGPTAFI 272
Db 188 VNLGQVVGWNTAAADNFQLSGGGQGFALPIGQMAIAGQIRSGGSPVTHIGPTAFI 247
Qy 273 GVVDNNGARVORVVGSAAPASLIGTGDVITAVDGAIPINSATAMADALNGHHHPGDVIS 332
Db 248 GVVDNNGARVORVVGSAAPASLIGTGDVITAVDGAIPINSATAMADALNGHHHPGDVIS 307
Qy 333 VNWQTSKGGTRTGNVTLAEGPPA 355
Db 308 VTWQTSKGGTRTGNVTLAEGPPA 330

RESULT 13
US-10-369-983-4
; Sequence 4, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR FILING DATE: 2003-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1010
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB-102F fusion
; OTHER INFORMATION: protein
US-10-369-983-4

Query Match          90.5%; Score 1631; DB 15; Length 1010;
Best Local Similarity 99.1%; Pred. No. 2e-120;
Matches 320; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 33 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVIVDPNGVVL 92
Db 8 APPALSQDRFADFPALPLDPSAMVAQVAVQVNNINIKLGYNNNAVAGTGVIVDPNGVVL 67
Qy 93 NNHVIAGATDINAFSVSGSGTGYVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 152
Db 68 NNHVIAGATDINAFSVSGSGTGYVDVVGVDRTQDVAVLQLRGAGLPSAAIGGGVAVGEP 127
Qy 153 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 212
Db 128 VVAMGNSGGGGTTPRAVPGRVVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPV 187
Qy 213 VNLGQVVGWNTAAADNFQLSGGGQGFALPIGQMAIAGQIRSGGSPVTHIGPTAFI 272
Db 188 VNLGQVVGWNTAAADNFQLSGGGQGFALPIGQMAIAGQIRSGGSPVTHIGPTAFI 247
Qy 273 GVVDNNGARVORVVGSAAPASLIGTGDVITAVDGAIPINSATAMADALNGHHHPGDVIS 332
Db 248 GVVDNNGARVORVVGSAAPASLIGTGDVITAVDGAIPINSATAMADALNGHHHPGDVIS 307
Qy 333 VNWQTSKGGTRTGNVTLAEGPPA 355
Db 308 VTWQTSKGGTRTGNVTLAEGPPA 330
```

Db 128 VVAMGNSGGGTPRAVPGRVVVALGQTVQASDSLTCABETLNGLIQFDAAIQPDAGGPV 187
QY 213 VNLGLGVGMNTAAADNFOLSGGGGFAIPICQAMAIAGQIRSGGGSPVTHIGTAFGL 272
Db 188 VNLGLGVGMNTAAADNFOLSGGGGFAIPICQAMAIAGQIRSGGGSPVTHIGTAFGL 247
QY 273 GVVDDNNGGARVQVRVVGSAASLGIQSTGDIIVAVDGAIPINSATAMADALNGHHPGDVIS 332
Db 248 GVVDDNNGGARVQVRVVGSAASLGIQSTGDIIVAVDGAIPINSATAMADALNGHHPGDVIS 307
QY 333 VNWQTKSGGTRTGNVTNLAEGPPA 355
Db 308 VNWQTKSGGTRTGNVTNLAEGPPA 330

RESULT 14

US-10-369-983-18
; Sequence 18, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85b)
US-10-369-983-18

Query Match 55.5%; Score 999.5; DB 15; Length 1016;
Best Local Similarity 84.1%; Pred. No. 26-70; Indels 15; Gaps 4;
Matches 207; Conservative 6; Mismatches 18
QY 28 APAQAAPPALSDQRFADFPALDPSAMVAQVAVPQVNNINTKLGYNNAVAGAGTGIVIDPN 87
Db 530 AAGDIAPPALSDQRFADFPALDPSAMVAQVAVPQVNNINTKLGYNNAVAGAGTGIVIDPN 589
QY 88 GVLTNNHVIAGATDINAFSVGSGGTGYGVVVDRTQDVAVLQIRGAGGLPSAIGGV 147
Db 590 GVLTNNHVIAGATDINAFSVGSGGTGYGVVVDRTQDVAVLQIRGAGGLPSAIGGV 649
QY 148 AVGEPVAMGNSGGGTPRAVPGRVVVALGQTVQASDSLTCABETLNGLIQFDAAIQPD 207
Db 650 AVGEPVAMGNSGGGTPRAVPGRVVVALGQTVQASDSLTCABETLNGLIQFDAAIQPD 709
QY 208 SGPVYVNGLGQVVGMMNTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGGSPVTHIGPT 267
Db 764 SPAVYL 769

RESULT 15

US-10-369-983-16
; Sequence 16, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff

; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB114F (MTB72F-MTCC#2)
US-10-369-983-16

Query Match 55.0%; Score 992; DB 15; Length 1154;
Best Local Similarity 63.8%; Pred. No. 9-28-70;
Matches 219; Conservative 17; Mismatches 59; Indels 48; Gaps 6;
QY 28 APAQAAPPALSDQRFADFPALDPSAMVAQVAVPQVNNINTKLGYNNAVAGAGTGIVIDPN 87
Db 530 AAGDIAPPALSDQRFADFPALDPSAMVAQVAVPQVNNINTKLGYNNAVAGAGTGIVIDPN 589
QY 88 GVLTNNHVIAGATDINAFSVGSGGTGYGVVVDRTQDVAVLQIRGAGGLPSAIGGV 147
Db 590 GVLTNNHVIAGATDINAFSVGSGGTGYGVVVDRTQDVAVLQIRGAGGLPSAIGGV 649
QY 148 AVGEPVAMGNSGGGTPRAVPGRVVVALGQTVQASDSLTCABETLNGLIQFDAAIQPD 207
Db 650 AVGEPVAMGNSGGGTPRAVPGRVVVALGQTVQASDSLTCABETLNGLIQFDAAIQPD 709
QY 208 SGPVYVNGLGQVVGMMNTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGGSPVTHIGPT 267
Db 710 SGPVYVNGLGQVVGMMNTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGGSPVTHIGPT 760
QY 268 AFLGLGVDDNNGGARVQVRVVGSAASLGIQSTGDIIVAVDGAIPICQAMAIAGQIRSGGGSPVTHIGPT 325
Db 761 ANDGV-----AAELTSAASVYSGVSVVSTLIVEPMMGPFAPAAAATAATPY 803
QY 326 HPGDVISVNW-----QTKSGGTRTGNVTNLAEGPPA 355
Db 804 -----VGLAATAALAKETATQAAAAAFAFGTAFAMTVPPS 839

Search completed: June 30, 2004, 17:14:43
Job time : 28.6821 secs

1. The first part of the document is a letter from the President of the United States to the Congress, dated January 3, 1862. It is a very long letter, and it contains a great deal of information about the state of the country at that time. It is a very important document, and it is one of the most interesting documents in the collection.

2. The second part of the document is a letter from the President of the United States to the Congress, dated January 3, 1862. It is a very long letter, and it contains a great deal of information about the state of the country at that time. It is a very important document, and it is one of the most interesting documents in the collection.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:31 ; Search time 9.27806 Seconds
(without alignments)
3680.509 Million cell updates/sec

Title: US-09-597-796c-4
Perfect score: 1802
Sequence: 1 MNSRRRLRWLLSVLAA.....QTKSGGTRTGNVTLAEGPPA 355
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.78:*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1792	99.4	355	2 F70983	probable serine pr
2	1306	72.5	361	2 S47170	hypothetical prote
3	1272.5	70.6	354	2 A87242	probable secreted
4	468.5	26.0	464	2 C70821	probable serine pr
5	460	25.5	382	2 H86930	probable secreted
6	460	25.5	452	2 T45448	probable serine pr
7	382	21.2	394	2 S74643	proteinase hhoA (E
8	372	20.6	407	2 A82150	serine proteinase
9	355.5	19.7	452	2 S77538	serine proteinase
10	354	19.6	433	2 H97139	htrA-like serine p
11	345	19.1	362	2 T35287	probable secreted
12	343	19.0	457	2 AG0433	proteinase (EC 3.4
13	341.5	19.0	514	2 A82591	periplasmic protei
14	335.5	18.6	429	2 A81894	serine proteinase
15	331	18.4	416	2 A82057	serine proteinase
16	331	18.4	441	2 E75357	probable periplasm
17	331	18.4	455	2 C91142	serine endoprotein
18	331	18.4	455	2 F85987	serine endoprotein
19	329.5	18.3	474	2 F83550	serine proteinase
20	329	18.3	398	2 B71284	probable periplasm
21	329	18.3	455	2 A80909	serine protease (E
22	328	18.2	355	1 J06082	trypsin-like prote
23	328	18.2	355	2 D91142	trypsin-like prote
24	327	18.1	352	2 E82307	trypsin-like prote
25	327	18.1	355	2 G85987	trypsin-like prote
26	326	18.1	455	2 J06031	trypsin-like prote
27	322	17.9	475	1 S15337	heat shock protein
28	319	17.7	475	2 AC0528	protease DO precu
29	317.5	17.6	474	2 S45229	proteinase DO (EC

ALIGNMENTS

RESULT 1

F70983
probable serine proteinase pepA - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002
C:Accession: F70983
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, N.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Scares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: F70983
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-355 <COL>
A:Cross-references: GB:Z96071; GB:AL123456; NID:G3242254; PIDN:CA809453.1; PID:G2181967
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: pepA
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; tryp

Query Match	99.4%;	Score 1792;	DB 2;	Length 355;
Best Local Similarity	99.4%;	Pred. No. 7.5e-98;		
Matches 353;	Conservative 0;	Mismatches 2;	Indels 0;	Gaps 0;
QY	1	MNSRRRSRLRWSWLLSVLAAVGLGLATAPAAQAAPPALSDRFPADFPALPLDPSAMVAQVA	60	
Db	1	MNSRRRSRLRWSWLLSVLAAVGLGLATAPAAQAAPPALSDRFPADFPALPLDPSAMVAQVG	60	
QY	61	PQVWVNTKLGYNNAVAGCTGIVDPNGVVLNNHVIAGATDINAFSVSGGGQTYGVDVVG	120	
Db	61	PQVWVNTKLGYNNAVAGCTGIVDPNGVVLNNHVIAGATDINAFSVSGGGQTYGVDVVG	120	
QY	121	YDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPFVWAMGNSGGQGTTPRAVPGRVVALGQTV	180	
Db	121	YDRTQDVAVLQLRGAGGLPSAAIGGVAVGEPFVWAMGNSGGQGTTPRAVPGRVVALGQTV	180	
QY	181	QASDSLTGAETLNGLIQFDAAIQPGDSGGPPVNGLGQTVVGNMTAASDNFQLSQSGQGGA	240	
Db	181	QASDSLTGAETLNGLIQFDAAIQPGDSGGPPVNGLGQTVVGNMTAASDNFQLSQSGQGGA	240	
QY	241	IPFGQMAIAGQIRSGGGSPTVHIQPTAFGLGLGVDDNNGGARVQRVVGSAFAASLGIST	300	
Db	241	IPFGQMAIAGQIRSGGGSPTVHIQPTAFGLGLGVDDNNGGARVQRVVGSAFAASLGIST	300	
QY	301	GDVITAVDCAPINSATAMADALNGHHHPGDVIVSNWQTSKGGTRTGNVTLAEGPPA	355	
Db	301	GDVITAVDCAPINSATAMADALNGHHHPGDVIVSNWQTSKGGTRTGNVTLAEGPPA	355	

Db 355 AIPVDAQRIADELISTGKA-----SHASLGQVQVNTDKDTLGAKIVEVVGAGAAAGV 408

QY 299 STGDVITAVDGPATNSATAMADALNGHHGPDVIVSNWQTKSGGTRTGNVTILAE 351

Db 409 PKGVVTKVDDREINSADALVAARSKAPGATVALTFQDPGSGSRVQVTLGK 461

RESULT 5

H86930

Probable secreted serine proteinase [imported] - Mycobacterium leprae

C:Species: Mycobacterium leprae

C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002

C/Accession: H86930

R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holt, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, M.A.; Rutherford, K.M.

Nature 409, 1007-1011, 2001

A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squares, R.

A:Title: Massive gene decay in the leprosy bacillus.

A:Reference number: A86909; PMID:21128732; PMID:11234002

A:Accession: H86930

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-382 <STO>

A:Cross-references: GB:AL450380; NID:g13092536; PIDN:CAC29684.1; GSPDB:GN00147

C:Genetics:

A:Gene: ML0176

C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

Query Match 25.5%; Score 460; DB 2; Length 382;

Best Local Similarity 36.4%; Pred. No. 5.3e-20;

Matches 130; Conservative 52; Mismatches 121; Indels 54; Gaps 12;

QY 22 GLGLATAPAAQAPALPSQDRFADFPALPLDPSAMVAQVA-----PQVNNINTKLGYNNAV 77

Db 50 GAGPVTGPAASVPAANM-----PSGSVEQVAVKVPVSMVMTDLGQSE-- 94

QY 78 AGTGVIDPNGVLTNNHVIAGA-----TDINAFVSGSGQTVGVVGYDR 123

Db 95 EGSVILSADGLIITNNHVAARSKAPGPGGGLSPKTTVTFF--DRTASTFVVGADP 151

QY 124 TDVAVLQRLGAGGLPSAAIGGG--VAVGEPVAVMNGSGGQGTTPRAVPGRVVAGLQTVQ 181

Db 152 TSDIADVVRVQISGLTPTMGSSADLRVGPVAVGSPGLAGT--VTSGIVSALNRPVS 209

QY 182 ASDSLTGAEETLNGLIQFDAAIQGDSGGPVVNGLGQVVGNTAA-----SDNFQLSQG 235

Db 210 TTGE-SGNQNTVLDAIQTDAAINPFGNSGGLVNMGGQLVGNVSAIATLGADSGDAQSGSI 268

QY 236 GQGAIPITGOAMATAGQIRSGGSPTHIGPTAFILGLGVNDNNGN-GARVQVVGSPAPAA 294

Db 269 GLGFAIPVDQAKRIADELISTG--KATH-----ASLGQVQVATDKTPGAKVMDVVAGGAAA 322

QY 295 SLGISTGDVITAVDGPATNSATAMADALNGHHGPDVIVSNWQTKSGGTRTGNVTILAE 351

Db 323 NAAVPGVVLTKVDDRLISSADALVAARSKAPGDKVSLTYQDQSGSRVQVTLGK 379

RESULT 6

T45448

Probable serine proteinase (EC 3.4.21.-) MLCB373.28 [similarity] - Mycobacterium leprae

C:Species: Mycobacterium leprae

C>Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 09-Dec-2002

C/Accession: T45448

R;James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, February 1999

A:Reference number: Z22967

A:Accession: T45448

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-452 <JMS>

A:Cross-references: EMBL:AL035500; PIDN:CAB36690.1

A:Experimental source: cosmid L373

Query Match 21.2%; Score 382; DB 2; Length 394;

Best Local Similarity 31.8%; Pred. No. 2e-15;

Matches 116; Conservative 50; Mismatches 117; Indels 84; Gaps 14;

QY 13 WLLSLVAALVGLGLATAP-----AQAAPPA-----LSQDRFADFPALPLDPSAMVA 57

Db 13 YLLAFAGVTAFTANLPHAVAAADDLPAPVITAQASVPLTSESFV-----AAAAS 63

QY 58 QVAPQVNNITK-----LQYNNAV-----GAGTGVIDPNGVV 90

Db 64 RSGFAVVRIDTETVWTRTDPILDDPFFOEFGRSPVPRRIAGQSGGFIIDNSGII 123

QY 91 LTNHVIAGTADINAFVSGSGQTVGVVGYDRQDVAVLQLRGAG-GLPSAAIG--GGV 147

C:Genetics:

A:Note: MLCB373.28

C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

C:Keywords: hydrolase; serine proteinase

F:182,224,305/Active site: His, Asp, Ser #status predicted

Query Match 25.5%; Score 460; DB 2; Length 452;

Best Local Similarity 36.4%; Pred. No. 6.3e-20;

Matches 130; Conservative 52; Mismatches 121; Indels 54; Gaps 12;

QY 22 GLGLATAPAAQAPALPSQDRFADFPALPLDPSAMVAQVA-----PQVNNINTKLGYNNAV 77

Db 120 GAGPVTGPAASVPAANM-----PSGSVEQVAVKVPVSMVMTDLGQSE-- 164

QY 78 AGTGVIDPNGVLTNNHVIAGA-----TDINAFVSGSGQTVGVVGYDR 123

Db 165 EGSVILSADGLIITNNHVAARSKAPGPGGGLSPKTTVTFF--DRTASTFVVGADP 221

QY 124 TDVAVLQRLGAGGLPSAAIGGG--VAVGEPVAVMNGSGGQGTTPRAVPGRVVAGLQTVQ 181

Db 222 TSDIADVVRVQISGLTPTMGSSADLRVGPVAVGSPGLAGT--VTSGIVSALNRPVS 279

QY 182 ASDSLTGAEETLNGLIQFDAAIQGDSGGPVVNGLGQVVGNTAA-----SDNFQLSQG 235

Db 280 TTGE-SGNQNTVLDAIQTDAAINPFGNSGGLVNMGGQLVGNVSAIATLGADSGDAQSGSI 338

QY 236 GQGAIPITGOAMATAGQIRSGGSPTHIGPTAFILGLGVNDNNGN-GARVQVVGSPAPAA 294

Db 339 GLGFAIPVDQAKRIADELISTG--KATH-----ASLGQVQVATDKTPGAKVMDVVAGGAAA 392

QY 295 SLGISTGDVITAVDGPATNSATAMADALNGHHGPDVIVSNWQTKSGGTRTGNVTILAE 351

Db 393 NAAVPGVVLTKVDDRLISSADALVAARSKAPGDKVSLTYQDQSGSRVQVTLGK 449

RESULT 7

S74643

proteinase hhoA (EC 3.4.-.-) - Synechocystis sp. (strain PCC 6803)

N:Alternate names: protein sll1679

C:Species: Synechocystis sp.

A:Variety: PCC 6803

C>Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Dec-2002

C/Accession: S74643

R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, N. A. Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis

S.

A:Reference number: S74322; PMID:97061201; PMID:8905231

A:Accession: S74643

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-394 <KAN>

A:Cross-references: EMBL:D90900; GB:AB001339; NID:g1651768; PIDN:BAAL6795.1; PID:g165186f

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C:Genetics:

A:Gene: hhoA

C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

C:Keywords: hydrolase; proteinase

Query Match 21.2%; Score 382; DB 2; Length 394;

Best Local Similarity 31.8%; Pred. No. 2e-15;

Matches 116; Conservative 50; Mismatches 117; Indels 84; Gaps 14;

QY 13 WLLSLVAALVGLGLATAP-----AQAAPPA-----LSQDRFADFPALPLDPSAMVA 57

Db 13 YLLAFAGVTAFTANLPHAVAAADDLPAPVITAQASVPLTSESFV-----AAAAS 63

QY 58 QVAPQVNNITK-----LQYNNAV-----GAGTGVIDPNGVV 90

Db 64 RSGFAVVRIDTETVWTRTDPILDDPFFOEFGRSPVPRRIAGQSGGFIIDNSGII 123

QY 91 LTNHVIAGTADINAFVSGSGQTVGVVGYDRQDVAVLQLRGAG-GLPSAAIG--GGV 147

```

Db      124 LTNHVVVDGASKV-VVTLRDGRFTDQVGRGTDEVTDLAVVKIEPQGSALFPVAPLGTSSNL 182
QY      148 AVCEPVVVMGSGGGTTPRAVPGRVVVG--QTVOASDSLTCABETLNGLLQFDAAATOP 205
Db      183 QVGDWIAVGNPVLGNT-----VTGLISTLGRSAAQAGIPDKRVEFTQTDAAINP 234
QY      206 GDSGGPVVNGLGQVGMNTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIG 265
Db      235 GNSGGPLLNARGVEVIGINTA---IRADATGIGFAIPIDQAKAIQNTLAAAGTVPHPIG 290
QY      266 PTAFGLGLGV---VDNNGN-----GARVQRVVGSAAPASLGISTGVDVITAVDGAP 311
Db      291 -VOMNITVDQAQNNRNPNPSPFIPEVDGILVMRVLPCTPAERAGIRRGDVIVAVDGT 349
QY      312 INSATAM 318
Db      350 ISDGAEL 356

RESULT 8
AG2150
serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AG2150
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:213595285; PMID:11759840
A:Accession: AG2150
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-407 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA074457.1; PID:g17131851; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: dir2758
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryg

```

```

Query Match      20.6%; Score 372; DB 2; Length 407;
Best Local Similarity 30.3%; Pred. No. 7,8e-15;
Matches 125; Conservative 59; Mismatches 141; Indels 88; Gaps 15;

QY      7 RSLR--WSLLSVLMAVGL---GLATAPQAAP-----PALSDRFPADFPALPLD 51
Db      10 RSIQLGTHLAIFFIGVVLTVSSRLVPSQAEPAPNPSTTGSAPELVAQRQSPTAAVIGN 69
QY      52 PS---AMVAQVAPVQVNTIKLYNNV-----GAGTG 81
Db      70 SSFVTAANVRVGSVAVRDTERTITRRVDPLEDPFRFRPFGEFGQQLPPEQVRGLGSG 129
QY      82 IVIDPNGVLLNNHVIAGATDINAFSVSGGTYGVVDVVDRTQDVAVLQRLGAGGLPSA 141
Db      130 FIIDKSGILTNHVVVDKADRVTV-RLKDGSRFDGKQGVDEVTDLAVVKNAGNSLFA 188
QY      142 AIG--GGVAVGCEPVVVMGSGGGTTPRAVPGRVVVG--QTVOASDSLTCABETLNGLI 197
Db      189 PLGSSNVVQGDWIAVGNPLGFNT-----VTGLIVSTLKRSAQGVITDKRLDFI 240
QY      198 QFDAAIQPDGSGPVVNGLGQVGMNTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGG 257
Db      241 QTDAAINPGNSGGPLLNDKGEVIGINTA---IRADAWGIGFAIPIDKAKAIATQLERDG 296
QY      258 GSPVTHIGPTAFILGV-----VDNNG-----NGARVQRVVGSAAPASLG 298
Db      297 --KVH-----PYLGVQWATITPELAQQNNIDPNSAFAPVWGLVIRVVPNSPAAV 350
QY      299 STGDVITAVDGAPINSATAMADALNGHHPGDVISVNMWQTKSGGTRTGNVTLAE 351
Db      351 RRGDVILLQVDGQAITTAELQNVVNSRLQALQVRLQ-RGNQTQQLSVTAE 402

```

```

RESULT 9
S77538
serine proteinase (EC 3.4.21.-) htrA - Synechocystis sp. (strain PCC 6803)
N:Alternate names: protein slr1204
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 09-Dec-2002
C:Accession: S77538
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;
O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda,
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A:Reference number: S74322; MUID:97061201; PMID:8905231
A:Accession: S77538
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-452 <KAN>
A:Cross-references: EMBL:D90905; GB:AB001339; NID:G1652360; PIDN:BA017385.1; PID:g165246;
A:Note: the nucleotide sequence was submitted to the EMBL data library, June 1996
C:Genetics:
A:Gene: htrA
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryps
C:Keywords: hydrolase; serine proteinase

Query Match      19.7%; Score 355.5; DB 2; Length 452;
Best Local Similarity 29.2%; Pred. No. 8.1e-14;
Matches 116; Conservative 65; Mismatches 129; Indels 87; Gaps 16;

QY      16 SVLAAVGLGLATAPQAAPALSDRFPADFPALPLDPSAMVAQV----APQVYNTIKLG 71
Db      84 SVISPLVTNQSIAPANESLATNLQSLRS-----PREPSNFVVDVVESTGPAVVRNAQKT 138
QY      72 YNNAV-----GAGTGIVDPNGVVLNNHVIAGATDIN 104
Db      139 VRSQVPAQFNDPFLRFPGSQMPNERNVQRGTSGFVSDNGKIFTNHVVVDGADEVT 198
QY      105 AFSVSGSGQTYGVVDVVDVVDRTQDVAVLQRLGAGGLPSAATGG--VAVGEPVVMGNSGGQ 162
Db      199 V-TLKDGSRFPGRVNGSDPSTDAVVVKIE-AGDLFTVALGSDHDLQVGWALAINP:GL 256
QY      163 GGTTPRAVPGRVVVGALQTVQASDSLTCABETLNGLLQFDAAATQPDGSGPVVNGLGQVGM 222
Db      257 DNT--VTTGILSATGR--RSAD--IGVPDKRVEFTQDAAINPGNSGGPFLNADQGVIGM 310
QY      223 NTAASDNFOLSGGGGFAIPICQAMAIAGQIRSGGSPVTHIGPTAFILGLGVVDNN--- 278
Db      311 NTAIONAQ---GIGFAIPINKAEIAQOLIATG--KVEH-----AYLGIQWVTWTPBLQ 360
QY      279 -----GNGARVQRVVGSAAPASLGISTGVDVITAVDGAPINSATAMADALNGH 325
Db      361 SOIROETGMNIPVDKGVVIMQVMPNSPAAIAKLEQGVLSLQSGQVQGVENAEQVQSLVGL 420
QY      326 HPGDVISVNMWQTKSGTGTG---NVTLAG-----PP 354
Db      421 AVGDEVEL-----GILNGQQQNTVTITGALPSAPP 451

RESULT 10
H97199
htrA-like serine protease (with PDZ domain) [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 09-Dec-2002
C:Accession: H97199
R:Nolling, J.; Bretton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: H97199
A>Status: preliminary

```

A:Molecule type: DNA
A:Residues: 1-433 <RUR>
A:Cross-references: GB:AE001437; PIDN:AAK80387.1; PID:G15025449; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics: CAC2433
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; tryP

Query Match 19.6%; Score 354; DB 2; Length 433;
Best Local Similarity 32.0%; Pred. No. 9.4e-14;
Matches 101; Conservative 51; Mismatches 110; Indels 54; Gaps 10;

QY 50 LDFSAMVAQVAPGVNINIKL-----GYNNAVAGAGTGIVDPNGVILT 92
DB 122 LTVSQIVKVPVAVGVSTKTITQNDFFSGSSNGSSSQEFGMSGLIIFNDCGYILT 181
QY 93 NNHVIAGATDINAFSGSGGTGVVVDVDRDQVAVLQIRGAGGLPSAAIGGVA---V 149
DB 182 NYHVIRKADKL-AVILNNKKEVSAKVNYDEANDIAVKTGSGFTVPGVAELGSSASLV 240
QY 150 GEPVAVNGSGGGGTTPRVPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQDSDG 209
DB 241 GDSVAIGNPLGKEFLGTVTGTVVSAVNRVAVSE---GQKQT---YIQTDAINFGNSG 294
QY 210 GPVYNGIGQVVGWMTAASDNFQISGGQ---GPAIPIGQAMALAGQIRSGGSPPTVHIGP 266
DB 295 GPLVNSFGQVVGNSA-----KISENGVEGHSIFIDTVKSKIQNLK-----P 339
QY 267 TAFILGLG--VVD-----NNGNGARVQVRVVGSAASLGISTGDTVITAVDGPAPNSATA 317
DB 340 ILMGISGEAVDKSTAEQHNIPQGVYIEIQDFSSAQKAGMQGVDTITKFGKKTSTSD 399
QY 318 MADALNGHHFGDVISV 333
DB 400 IDSISKXNSGDTVQV 415

RESULT 11
T35287
probable secreted proteinase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 09-Dec-2002
C:Accession: T35287
R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: 221574
A:Accession: T35287
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-362 <SBE>
A:Cross-references: EMBL:AL096872; PIDN:CAB51255.1; GSPDB:GN000070; SC0DB:SCSF7.30
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SC0DB:SCSF7.30
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; tryP

Query Match 19.1%; Score 345; DB 2; Length 362;
Best Local Similarity 32.6%; Pred. No. 2.6e-13;
Matches 125; Conservative 49; Mismatches 154; Indels 56; Gaps 14;

QY 1 MSNRRRLRNSWLLSVLAAGVLGATA-----PAQAPALSDQR 41
DB 1 MDTSRTRLRLAPVAVPACV-LLLATGCSAGAGTDRGSGSAREGDTAQAAAPRAASEL 59
QY 42 FADFPALPLDFSAMVAQVAPGVNINIKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGAT 101
DB 60 EADY-----ERVIKDVLPSVVIQA-----GDSLGSGVYVDKGHVTVNAHVVG--- 103
QY 102 DINAFSVSGSGQTYG---VDVYGYDRDQVAVLQL-RGAGGLPSA--AIGGGVAVGEPVWA 155
DB 104 DAKSPRVTTARTEGALTAKLVSSYPEQDLAVIKLDKVPENRAARFADSAKVEVGQIVLA 163
QY 156 MNSGCGGCTPRVPRVVALGQTVQASDSLTGAETLNGLIQFDAAIQDSDGSPVYNG 215

DB 164 MGSPLGLSSS--VTQGVISATGRVITSGSGGTGATIANVQVTSAAIPGNSGALVNL 221
QY 216 LGQVVGWMTAASDNFQISQG---GQGPALPIGQAMALAGQIRSG---GSPTVHIGPTAF 269
DB 222 DGQVIGIPTLAATDPGLGDSAAAPGIGFPAIPASWVTTVAGQIVRDKGKVTDSGRAALGITA- 280
QY 270 LGLGWDDNNGN--GARVQVRVVGSAASLGISTGDTVITAVDGPAPINSATAMADALNGHHP 327
DB 281 --RTVVDSDYRPAAGAAVVEVSDGGAADAGLRPGDVLVKLGDTDTITTSISEALASMRP 338
QY 328 GDVISVNMQTKSGGTRTGNVTLAE 351
DB 339 GDRTKVTV-TRDGKEHTAEVTLGE 361

RESULT 12
AG0433
proteinase (EC 3.4.21.-) [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 27-Nov-2001
C:Accession: AG0433
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.;
deno-Farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
Hill, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, B.
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AG0433
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-457 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC92795.1; PID:G15981488; GSPDB:GN00175
C:Genetics:
A:Gene: degQ
C:Superfamily: Helicobacter serine proteinase
C:Keywords: hydrolase; serine proteinase

Query Match 19.0%; Score 343; DB 2; Length 457;
Best Local Similarity 31.2%; Pred. No. 4.4e-13;
Matches 122; Conservative 53; Mismatches 126; Indels 90; Gaps 18;

QY 12 SWLLSVLA-AVGLGLATAP---AQAAPPALSDQRADFPALPLDPSAMVAQVAPGVNIN- 66
DB 5 SLLLSALATSVGLGLASVPMVSAALPAVAGQ---TLPSL-----APMLEKVLPAVSVH 57
QY 67 -----NTKLGYNNA---VGAGTGIVID-PNGVVLTNHVIAGAT 101
DB 58 VSGSQAQQQLPEEFKFFGPNAPSGKSSRRPFGLGSGVINAEGYILTNNHVINAD 117
QY 102 DINAFSVSGQTYGVVDVYGYDRDQVAVLQIRGAGGLPSAAI--GGVAVGEPVAVAGNS 159
DB 118 KIRV-QLNDGREYDAKLLGRDEQTDIALQLTDAKNLTAIKTADSDNRVGFVAVAGNP 176
QY 160 GSGGTPRAVPRVVALGQTVQASDSLTGAE--ETLNGLIQFDAAIQDSDGSPVYNGLQ 218
DB 177 FGLGQT--ATSGLIISLGR-----SGLNLEGLENFQTDASINRNGSGGALVNLDE 226
QY 219 VVGWMTAASDNFQISQGG---GPAIPIGQAMATA-----GQIRSG-----GSPTVH 263
DB 227 LIGINTAI-----LAPCGGNIGIGFPAIPASWVTTVAGQIVRDKGKVTDSGRAALGITA- 281
QY 264 IGTAFLGLGVDDNNGNGARVQVRVVGSAASLGISTGDTVITAVDGPAPINSATAMADALN 323
DB 282 IAKAFNI-----DAQRGAPVSEVLPKSAAKAGIKPGDVLISVDGKKLSSFAELRAKVG 335
QY 324 GHHPGDIVISVNMQTKSGGTRTGNVTLAE 354
DB 336 TTGFGTKIKIG-----LLREGKP 353

RESULT 13
A82581

periplasmic proteinase XF2241 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
R:Accession: AB2581
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequences
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: AB2515; MUID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: AB2581
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-514 <SIM>
A:Cross-references: GB:AE004037; GB:AE003849; NID:g9107394; PIDN:AAF85040.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpon, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvares, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir
M.; Tshako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
A:Genetics:
A:Gene: Xf2241
A:Superfamily: Helicobacter serine proteinase
Query Match 19.0%; Score 341.5; DB 2; Length 514;
Best Local Similarity 31.6%; Pred. No. 6.1e-13;
Matches 123; Conservative 55; Mismatches 128; Indels 83; Gaps 17;
QY 3 NSRRSLRWMLLSV---LAAVGLGLATAPAAAP--PALSDQDFADFPALDEL-----DP 52
Db 2 NSRIHT-RCFGLLAITLPLAACGAGHNSPTTAPSTPIVTP---STTPAPQLVAGLPDF 57
QY 53 SAMVAQAPQVNVNT-----KLGVNAV----- 76
Db 58 TQVVDQVGVNVNITVTRKKVGRGIPLDNIDIPFRFFFGDFQMPNQPRGGQDE 117
QY 77 -----GAGTGIVDPNGVLTNNHVIAGATDINAFSGQTYGVVDVDRDQDVAVL 130
Db 118 GGIAGRGMGSGFIISKGYILTNHVIAGSEV-TIKLTDREPKAKIIGSDQYDVALL 176
QY 131 QLRGAGGLPSAAGGVAV--GEPVAMNGSGGGTTPRVPVGVVAGLQTVQASDGLTG 188
Db 177 KI-DANKNLTPTVRIGSSSLKGGQWVAIGSPFLDHSVTA--GIVSALGRS-----TS 226
QY 189 AEETLNGLIQFDAIQTQDSGGPVVNGLGQVGVNMTAASNFQLSQG--GQGFAPITGQA 246
Db 227 DQQRVVPITQDVPIQNGSGGPIILNTRGEVIGIN---SQIFSASGVMGIFSAPINLA 283
QY 247 MAIAGQIRSG-----GSPTVHIGTAFJ---GLGVNNGNGARVQVVGSAAPASLGIS 299
Db 284 INAAEQIRKTKQVMSRLMGVEIGIPIDAKAQLGLPDS--RGALVNNIPPHSPAAKAGIE 341
QY 300 TGDVITAVDGPINSATAMALNGHHPG 328
Db 342 VGDVIRSVNGKVISFSFDLPPLIGMPPG 370
RESULT 14
AD1894
serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AD1894

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AD1894
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-429 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA72659.1; PID:g17130047; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2008
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; tryptophan
Query Match 18.6%; Score 335.5; DB 2; Length 429;
Best Local Similarity 29.2%; Pred. No. 1.1e-12;
Matches 110; Conservative 59; Mismatches 137; Indels 71; Gaps 14;
QY 18 LAAVGLGLATAPAAAPALSDQDFADFPALDELPSAMVAQVAPQVNVNITKLGVNAV- 76
Db 67 LAQKASDLAVSRVDAAPP--LGNTDQNF-----VTQVVRVGPVAVRISARTVTSRLP 119
QY 77 -----GAGTGIVDPNGVLTNNHVIAGATDINAFSGVSG 111
Db 120 AEFNDPFFERFFSGSLPQQQERVQRTGSGFLISADGSLTNAHVVDGADTVRVI-LKDG 178
QY 112 QTVGVVDVDRDQDVAVLQLRGAGGLPSAALTGGG--VAVGPPVAMNGSGGGTTPRAV 169
Db 179 RSFQKVLGTDLNLTDAVAVKIQ-ANNLPTLTGVNSDQLQFGQWAIAGNPLGLDNT--VT 235
QY 170 PGRVWALGOTVQASDLSLTCAEETLNGLIQFDAIQTQDSGGPVVNGLGQVGVNMTAASDN 229
Db 236 TGLISATGPT---SNQIGAPKRVETIOTDAINPNSGGPLINRGEVIGNATIIQG 291
QY 230 FQLSQGGQGFAPITGQAMAIAGQIRSGGGSPVHIGTAFGL-----GVV 275
Db 292 AQ-----GLGFAIPITQAIRISNQLIATGKVPYILG-IQWVGLTPQIKONINSDENSGLT 346
QY 276 DNNNGARVQVVGSAAPASLGISTGDDVITAVDGPINSATAMALNGHHPGDIIVSN- 334
Db 347 VDRKGLVVRVVPNSPAAAGLRAGDVITQKLNGQAVTDASNVQRAVENAQVGGQLQLEL 406
QY 335 WQTKSGGTRTG-NVTILA 350
Db 407 W-----RNGRNINLA 416
RESULT 15
AB2057
serine proteinase [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AB2057
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB2057
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-416 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA73707.1; PID:g17131098; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2008
C:Superfamily: Escherichia coli trypsin-like proteinase degS; GLGF domain homology; tryptophan
Query Match 18.4%; Score 331; DB 2; Length 416;
Best Local Similarity 31.5%; Pred. No. 2e-12;
Matches 92; Conservative 57; Mismatches 115; Indels 28; Gaps 10;

QY	77	GAGTGIIVDPNGVLTNNHVIAGTDINAFVSGSQTYGVDVGVGYDRTQDVAVLQLRGAG	136
Db	133	SGSGFIISSSGQIITNAHVVDGDEVTV-TLKGRSPGKVLGEDPVDVAVIQI-NAN	190
QY	137	GLPSAAIGGVAV--GEPVAMNSGGQGTTPRAVPGRVVALGQTVQASDSLTAETLN	194
Db	191	NLPTVAVGNSEVLQFGEAVIAIGNPLGLNNS-VTSGIISATGR---SGSDIGASDKRV	244
QY	195	GLIQDAAIQDGGPVGNGLVGVGMNTAASDNFQLSQGGQGEAIPIGQAMAIAGQIR	254
Db	245	DYLOTDAAINPGNSGGLNARGQVIGMNTAIQGAQ---GLGFAIPINTVQKVQSQELI	300
QY	255	SGGSGPTVHIG-PTAFLGLGVVD--NNGNGARVQ-----RVVGSAPAASLGISTG	301
Db	301	TQKVDHPYLGVMATLTPQVKERINEREGDRINITADRGVLLVRIVFGSPAANAGLRPG	360
QY	302	DVITAVDGAIPINSATAMADALNGHHPGDVISVNWOTKSGGTGTGNTLAEGP	353
Db	361	DIITOSINNOQVTTVEQVKIVENSQIQPLQIQIE-RNGQTTQVNVSPAPLP	411

Search completed: June 30, 2004, 16:53:57
Job time : 13.2781 secs

1. The first part of the document is a letter from the President of the United States to the Congress, dated January 3, 1862. It is a very long letter, and it contains a great deal of information about the state of the country at that time. The President talks about the war, the economy, and the future of the nation. He also talks about the role of the President and the Congress.

2. The second part of the document is a letter from the Secretary of the Treasury to the President, dated January 10, 1862. It is a very short letter, and it contains a great deal of information about the state of the Treasury at that time. The Secretary talks about the war, the economy, and the future of the nation.

3. The third part of the document is a letter from the Secretary of the Treasury to the Congress, dated January 10, 1862. It is a very short letter, and it contains a great deal of information about the state of the Treasury at that time. The Secretary talks about the war, the economy, and the future of the nation.

4. The fourth part of the document is a letter from the Secretary of the Treasury to the President, dated January 10, 1862. It is a very short letter, and it contains a great deal of information about the state of the Treasury at that time. The Secretary talks about the war, the economy, and the future of the nation.

5. The fifth part of the document is a letter from the Secretary of the Treasury to the Congress, dated January 10, 1862. It is a very short letter, and it contains a great deal of information about the state of the Treasury at that time. The Secretary talks about the war, the economy, and the future of the nation.

6. The sixth part of the document is a letter from the Secretary of the Treasury to the President, dated January 10, 1862. It is a very short letter, and it contains a great deal of information about the state of the Treasury at that time. The Secretary talks about the war, the economy, and the future of the nation.

7. The seventh part of the document is a letter from the Secretary of the Treasury to the Congress, dated January 10, 1862. It is a very short letter, and it contains a great deal of information about the state of the Treasury at that time. The Secretary talks about the war, the economy, and the future of the nation.

8. The eighth part of the document is a letter from the Secretary of the Treasury to the President, dated January 10, 1862. It is a very short letter, and it contains a great deal of information about the state of the Treasury at that time. The Secretary talks about the war, the economy, and the future of the nation.

9. The ninth part of the document is a letter from the Secretary of the Treasury to the Congress, dated January 10, 1862. It is a very short letter, and it contains a great deal of information about the state of the Treasury at that time. The Secretary talks about the war, the economy, and the future of the nation.

10. The tenth part of the document is a letter from the Secretary of the Treasury to the President, dated January 10, 1862. It is a very short letter, and it contains a great deal of information about the state of the Treasury at that time. The Secretary talks about the war, the economy, and the future of the nation.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:42:21 ; Search time 5.77978 Seconds
(without alignments)
3198.204 Million cell updates/sec

Title: US-09-597-796c-4
Perfect score: 1802
Sequence: 1 MSNRRRLRWMLLSVLA.....QTKSGGTRTCNVTLAEPPA 355

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	350.5	19.5	458	1	YVTA_BACSU
2	328	18.2	458	1	DEGS_ECOLI
3	326	18.1	455	1	DEGQ_ECOLI
4	322	17.9	475	1	DEGP_SALTY
5	321	17.8	413	1	HTRA_LACHE
6	317.5	17.6	474	1	DEGP_ECOLI
7	310	17.2	408	1	HTRA_LACLA
8	306	17.0	513	1	DEGP_BRUME
9	305	16.9	513	1	DEGP_BRUSU
10	295.5	16.4	449	1	HTRA_BACSU
11	295	16.4	466	1	HTRA_HAEIN
12	291	16.1	504	1	DEGP_RHIME
13	290	16.1	503	1	DEGP_BARHE
14	289.5	16.1	437	1	DEGL_ARATH
15	288	16.0	497	1	DEGP_CHLTR
16	286.5	15.9	497	1	DEGP_CHLMU
17	286	15.9	478	1	DEGP_EUCAI
18	283	15.7	508	1	DEGP_RICCN
19	282	15.6	488	1	DEGP_CHLPN
20	276.5	15.3	400	1	YVXA_BACSU
21	276	15.3	448	1	HRA3_ARATH
22	272.5	15.1	478	1	DEGP_EUCAP
23	263	14.6	458	1	HRA2_HUMAN
24	258	14.3	513	1	DEGP_RICPR
25	254	14.1	458	1	HRA2_MOUSE
26	244.5	13.6	340	1	DEGS_HAEIN
27	238.5	13.2	453	1	HRA3_HUMAN
28	235	13.0	480	1	HRA1_HUMAN
29	231.5	12.8	476	1	HRA4_HUMAN
30	225.5	12.5	460	1	HRA3_MOUSE
31	225	12.5	480	1	HRA1_MOUSE
32	178.5	9.9	321	1	SEPA_ARATH
33	173	9.6	630	1	Y4BJ_RHISN

RESULT 1

YVTA_BACSU					
ID	YVTA_BACSU	STANDARD;	PRT;	458 AA.	
AC	Q9R9I1; O35021; O35039;				
DT	28-FEB-2003 (Rel. 41, Created)				
DT	28-FEB-2003 (Rel. 41, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DE	Probable serine protease Yvta (EC 3.4.21.-).				
GN	YVTA OR YVTB OR BSU33000.				
OS	Bacillus subtilis.				
OC	Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.				
OX	NCBI_TaxID=1423;				
RN	[1]_TaxID=1423;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168;				
RX	MEDLINE=20158875; PubMed=10692364;				
RA	Noone D., Howell A., Devine K.M.;				
RT	"Expression of ykdA, encoding a Bacillus subtilis homologue of Htra,				
RT	is heat shock inducible and negatively autoregulated.";				
RL	J. Bacteriol. 182:1592-1599 (2000).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=98015415; PubMed=9353931;				
RA	Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;				
RT	"Sequencing of regions downstream of addA (98 degrees) and citG (289				
RT	degrees) in Bacillus subtilis.";				
RL	Microbiology 143:3305-3308 (1997).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=168;				
RX	MEDLINE=98044033; PubMed=9384377;				
RA	Kunst F., Ogaawara N., Moszer I., Albertini A.M., Alloni G.,				
RA	Azevedo V., Bartero M.G., Bessieres P., Bolotin A., Borchert S.,				
RA	Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,				
RA	Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,				
RA	Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,				
RA	Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,				
RA	Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,				
RA	Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,				
RA	Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,				
RA	Guiseppi H., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,				
RA	Hilbert G., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,				
RA	Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,				
RA	Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,				
RA	Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,				
RA	Lee S.M., Levine A., Liu H., Masuda S., Mavel C., Medigue C.,				
RA	Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,				
RA	Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,				
RA	Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,				
RA	Priegan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,				
RA	Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadaie Y.,				
RA	Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,				
RA	Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,				
RA	Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,				
RA	Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,				
RA	Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,				

34	162.5	9.0	457	1	PRTC_STRGR	P52320 streptomyc
35	158.5	8.8	747	1	ELS_BOVIN	P04985 bos taurus
36	153	8.5	801	1	Y747_MYCTU	O53810 mycobacteri
37	150.5	8.4	515	1	Y740_MYCTU	O50594 mycobacteri
38	147	8.2	299	1	PRTB_STRGR	P00777 streptomyc
39	146.5	8.1	864	1	ELS_RAT	O99372 rattus norv
40	145	8.0	914	1	WA22_MYCTU	O06794 mycobacteri
41	142.5	7.9	957	1	Y278_MYCTU	P64767 mycobacteri
42	141.5	7.9	710	1	F1B1_ADEGL	Q64761 avian adeno
43	141	7.8	2021	1	OMPA_RICCN	Q52657 rickettsia
44	139.5	7.7	778	1	YQ34_MYCTU	P71933 mycobacteri
45	139.5	7.7	2364	1	PGCA_BOVIN	F13608 bos taurus

ALIGNMENTS

RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 subtilis".
 RL Nature 390:249-256(1997).
 RN [4]
 RN TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX MEDLINE=20576168; PubMed=11133960;
 RA Noone D., Howell A., Collier R., Devine K.M.;
 RT "Ykda and YvtA, HraA-like serine proteases in *Bacillus subtilis*,
 RT engage in negative autoregulation and reciprocal cross-regulation of
 RT ykda and yvtA gene expression.";
 RL J. Bacteriol. 183:654-663(2001).
 CC -!- FUNCTION: May be involved in processing, maturation, or secretion
 CC of extracellular enzymes.
 CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
 CC -!- INDUCTION: Induced by heat shock during exponential growth and by
 CC heterologous amylases at the transition phase of the growth cycle.
 CC Negatively regulates its own expression.
 CC -!- MISCELLANEOUS: Inactivation results in compensating overexpression
 CC of htrA, especially during stress conditions.
 CC -!- SIMILARITY: Belongs to peptidase family S2C.
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC -!- CAUTION: Ref.2 and Ref.3 sequences differ from that shown due to
 CC frameshifts in positions 87 and 246 that produce two separate
 CC ORFs.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF188296; AAF03153.1; -;
 CC EMBL; Z93941; CAB07968.1; ALT_FRAME.
 CC EMBL; Z93941; CAB07969.1; ALT_FRAME.
 CC EMBL; Z93120; CAB15230.1; ALT_FRAME.
 CC Subtilisin; BG4155; yvtA.
 CC InterPro; IPR009003; Cys_Ser_trypsin.
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR001254; Peptidase S1.
 CC InterPro; IPR001940; Peptidase_S1C.
 CC Pfam; PF00595; PDZ; 1.
 CC Pfam; PF00889; trypsin; 1.
 CC PRINTS; PRO0834; PROTEASES2C.
 CC SMART; SM00228; PDZ; 1.
 CC PROSITE; PS0106; PDZ; 1.
 CC Hydrolase; Protease; Serine protease; Heat shock; Transmembrane;
 KW Complete proteome.
 FT DOMAIN 1 71 CYTOSOLIC (POTENTIAL).
 FT TRANSMEM 72 92 POTENTIAL.
 FT DOMAIN 93 458 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 356 440 PDZ.
 FT ACT SITE 187 187 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 217 217 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 298 298 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 458 AA; 48717 MW; 77551045865A5CD CRC64;
 Query Match 19.5%; Score 350.5; DB 1; Length 458;
 Best Local Similarity 27.6%; Pred. No. 8.3e-13;
 Matches 111; Conservative 67; Mismatches 153; Indels 71; Gaps 16;
 QY 5 RRSRLWSWLSVLA--GGLATAP-----AQAAPPALSDRFADFPALPLDPS 53
 DB 66 KEKKRAAWLSPILGIGGGLMGLIAPYLPDQNTATETASANKVQSDNFTAPTNA 125
 QY 54 A---MVAQVAPVYNI-NTKLGVYNAV-----GAGTGIVI---DPNGVLTN 93
 DB 126 SNIAADWVEDLEPTIVGISNIQTQNNFTGCGGSSSESCTGSGVIFPKDSKRAYITN 185

QY 94 NHVIAGATDINAFSVSGGOTGVVGVYDRTQVAVLQLRGAGLPSPAIGGG--VAVGE 151
 DB 186 NHVVEGANKLTV-TLYNGETAKLVGSDTITDLAVLEISGKNVKKVAFSGDSQLRTGE 244
 QY 152 PTVAMNSGGOGTTPRAVGRVVALGQTQVQASDLSLTGAETLNGLIQFPAALPGDGGGP 211
 DB 245 KVIAIGNPLGGQFSGVTQGIISGLNRTIDV-DITQGTVM--NVLTQDALNPGNSGP 301
 QY 212 VVNLGQVGMVMTAANDNFQSGG---QGFAIPQQAIAQIRSGGSGSPVTHIGPTA 268
 DB 302 LINASGVIGIN-----SLKVSSESVESLGFALPSNDVPIVDQLLQNKV-----DRP 350
 QY 269 FLGLGVVDNN-----GNGARVQRVVGSAFASASLGISTGVDVITAVDGA 310
 DB 351 FLGVQIMDSQVPEYQENTLGLFGDQLGKGVVYKVEVQANSFAEKAGIKSEDIIVKLNGK 410
 QY 311 PINSATAMADAL-NGHHPGDVIVSNWQTKSGGTRTGNVTLAE 351
 DB 411 DVSESSADIRQLLYKDLKVGDKTIIQ-VLEKGTIKILNATLTK 451
 RESULT 2
 DEGS ECOLI
 ID DEGS ECOLI STANDARD; PRT; 355 AA.
 AC P31137;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 GN Protease degs precursor (EC 3.4.21.-).
 DE DEGS OR HOB OR HHRH OR B3235 OR Z4594 OR ECS4108.
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID=562, 83334;
 RN [1]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RA Bass S., Gu Q., Goddard A.;
 RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=96165272; PubMed=8576051;
 RA Waller P.R., Sauer R.T.;
 RT "Characterization of degQ and degS, Escherichia coli genes encoding
 RT homologs of the DegP protease.";
 RL J. Bacteriol. 178:1146-1153(1996).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RN SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
 RL Nature 409:529-533(2001).
 RN [5]
 RN SEQUENCE FROM N.A.

RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
Kihara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic *Escherichia coli*
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP PRELIMINARY SEQUENCE OF 82-355 FROM N.A.
RX MEDLINE=98105815; PubMed=3322223;
RA Vogel R.F., Entian K.-D., Mecke D.;
RT "Cloning and sequence of the mdh structural gene of *Escherichia coli*
coding for malate dehydrogenase.";
RL Arch. Microbiol. 149:36-42(1987).
RN [7]
RP IDENTIFICATION.
RA Bazan J.F., Fletterick R.J.;
RT "Structural and catalytic models of trypsin-like viral proteases.";
RL Semin. Virol. 1:311-322(1990).
CC -!- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U15661; AAC43993.1; -
DR EMBL; U32495; AAC44006.1; -
DR EMBL; U18997; AAA58037.1; -
DR EMBL; AE000402; AAC76267.1; -
DR EMBL; AE005551; AAG58363.1; -
DR EMBL; AP002564; BAB37531.1; -
DR EMBL; M24777; -; NOT_ANNOTATED_CDS.
DR PIR; D91142; D91142.
DR PIR; JC6052; JC6052.
DR MEROPS; S01.275; -
DR EcoGene; EG11652; hhoB.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR PRINTS; PR00834; PROTEASES2C.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00595; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
DR Hydrolase; Serine protease; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 28
FT CHAIN 29 355
FT DOMAIN 281 326
FT ACT_SITE 96 96
FT ACT_SITE 126 126
FT ACT_SITE 201 201
FT ACT_SITE 253 253
FT CONFLICT 307 307
FT CONFLICT 307 307
FT CONFLICT 307 307
SQ SEQUENCE 355 AA; 37581 MW; D091B4D65E8FE1CC CRC64;
Query Match 18.2%; Score 328; DB 1; Length 355;
Best Local Similarity 30.7%; Pred. No. 1.1e-11;
Matches 110; Conservative 47; Mismatches 135; Indels 66; Gaps 13;
OY 15 LSVLAAVGLG-----LATAPAAQAPALSQDFADFPALPDPSAMVAQVAPQVNVIN 67
DB 3 VKLLRSVAIGHIVGAILLVAMPSSINPLSTPQDFSTDETPASYNLAVERAAPAVNVV- 61
OY 68 TKLGYNNAVGA-----GTGIVIDPNGVVLNNHVIAGATDINAFSVSGQTYGV 116

Db 62 ----YNRGLTNSHNQLRLTSGVINDQRYITNKHVINDAQI-IVALQDGRVFEA 116
OY 117 DVGVDRDTQDVAVLQLRAGGLPSAAICGGVA--VGEFVVMGNSGGGGTTPRAVPGRVV 174
Db 117 LLVGSDSLTDLAVLKINATGGLPTIPINARRVPHIGDVVLAIGNP-----Y 162
OY 175 ALGQTV-CASDSLTAABETLN-----GLIQDAAIQDSCGPPVVGVLGQVVMNTAASD 228
Db 163 NLGQITQGIISATG-RIGLNTGTQNFQPLQDASINHGNSGALVNSLGELMGLNTLSFD 221
OY 229 --NFOLSQGGGGAIP-----IGQMAIAGQIRSGGSGFTVHIGTAFILGLGVV 275
Db 222 KSNDETPEGIGFAIPFQATKIMDKLRDGRVIRYIGIGR----EIAPLHAQGGGI- 276
OY 276 DNNNGGARVQVWGSAPASISIGTGDVITAVDGAIPNSATAMADALNGHHPGDVISV 333
Db 277 -DQLGIVVNEVSPDGPAAAGIQVNDLIIISVDNKPALSALETWDQVAIRPGSVIPV 333
RESULT 3
DEQG_ECOLI STANDARD; PRT; 455 AA.
AC P30999;
DT 01-FEB-1995 (Rel. 31; Created)
DT 01-FEB-1995 (Rel. 31; Last sequence update)
DT 28-FEB-2003 (Rel. 41; Last annotation update)
DE Protease deqQ precursor (EC 3.4.21.-).
GN DEQG OR HHOA OR B3234.
OS *Escherichia coli*.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; *Escherichia*.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Bass S., Gu Q., Goddard A.; EMBL/GenBank/DBJ databases.
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-32.
RC STRAIN=K12 / W3110;
RX MEDLINE=96165272; PubMed=8576051;
RA Waller P.R., Sauer R.T.;
RT "Characterization of deqQ and deqS, *Escherichia coli* genes encoding
homologs of the deq protease.";
RJ J. Bacteriol. 178:1146-1153(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RJ Science 277:1232-1244(1997).
CC -!- FUNCTION: PROTEASE WITH A SHARED SPECIFICITY WITH DEGP/DESP.
CC -!- SUBCELLULAR LOCATION: Periplasmic.
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U15661; AAC43993.1; -
DR EMBL; U32495; AAC44006.1; -
DR EMBL; U18997; AAA58036.1; -
DR EMBL; AE000402; AAC76266.1; -

```

DR PIR: JC6051; JC6051.
DR MEROPS; S01.274; ---
DR SWISS-2DPAGE; P39059; COLI.
DR EcoGene; EGI2612; degQ.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
KW Hydrolase; Serine protease; Periplasmic; Repeat; Signal;
KW Complete proteome.
FT SIGNAL 1 27
FT CHAIN 28 455
FT DOMAIN 258 349
FT DOMAIN 355 447
FT ACT_SITE 109 109
FT ACT_SITE 139 139
FT ACT_SITE 214 214
FT ACT_SITE 214 214
SQ SEQUENCE 455 AA; 47205 MW; 6A090F93AC021C83 CRC64;

Query Match
Best Local Similarity 18.1%; Score 326; DB 1; Length 455;
Matches 119; Conservative 61; Mismatches 117; Indels 94; Gaps 19;

QY 14 LLSVLA-AVGLGL-ATAPAAAPPALSDQRFADFPALPDSAMVAQVAPQVNT- 68
DQ |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DQ 7 LLSALALSGLTSLASFQAVASIPQGVADQ-APLPSL-----APMLEKVLPAVSVRVEGT 61
QY 69 -----KLGNNAV-----GAGTGVIDPN-GVLTNNHVIAGATDINAFS 107
DQ |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DQ 62 ASQGGKIPFEKFKFGDDLPQAPQPFGLSGVLIINASKYVLTNNHVINQAQKI-SIQ 120
QY 108 VSGGQTYGVVGVYDRTQDVAVLQIRGAGGLPSAIGGG--VAVGEPVVMNGSGGGGT 165
DQ |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DQ 121 LNDGREFDAKLGSDQSDQDALLQTONPSKLTQIAADSDKLRVGDFAVAVGNPFLGQT 180
QY 166 PRAVGRVVALQQTVCASDSLTGAE-ETLNGLIQFDAAIQPDGSGPVNGLGVGQVNT 224
DQ |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DQ 181 --ATSGIVSALGR-----SGLNLEGFNIQTDAISINRENGGALLNGLIGINT 230
QY 225 AASDNFQLSQGGQ-----GFAIPIGQMAIA-----QOIRSGGSPVTHIGFTAFGLGV 274
DQ |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DQ 231 AI-----LAPGGSGVIGFAIPSNMARTLAQQLIDFGEIKRG-----LLGIKG 273
QY 275 VDNNGN-----CARVQVVGSAAPASLSIGTGDVITAVDVGAPINSATAMADLN 323
DQ |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DQ 274 TEMSADIKAFNLDVQRAVFSEVLPGSGSAGVAGKADGDIITSLNGKPLNSFAELSRIA 333
QY 324 GHHPGDVIVNMQTKSGGTRTG-----NVTL 349
DQ |||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
DQ 334 TTEPGT-----KVKLGLENGKPLEVEVTL 358

RESULT 4
DEGP SALTY STANDARD; PRT; 475 AA.
AC P26982;
DT 01-AUG-1992 (Rel. 23, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Protease do precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR PTD OR STMO209.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=602;
RN [1]_TaxID=602;
RP SEQUENCE FROM N.A.
RC STRAIN=CS;

```

```
Db 128 VTNNVVNDASVTKV-QLSDGRKFDKVKVQKPRSDIALIQONPKNLTAIKLADSDALR 186
Qy 149 VGBPVVAMGNSGGGGTPRAVPRVVALGQTVQASDSLTGAE-ETLNGLIQFDAAIQPGD 207
Db 187 VGDYTVAINGPFGLGET--VTSGIVSALGR-----SGLNVENYENFIQTDAALNRGN 236
Qy 208 SGBPVVNLGQVVMNTA--ASDNFQSQGGQFAIP-----IGQAMAIAGQIRSGGS 259
Db 237 SGGALVNLNGELIGINTAILAPDGGNI---GIGFAIPSNMVKNTSQMVEYGGVXRG--- 290
Qy 260 PTVHIGPTAFLGLGVNDNNGN-----GARVORVWGSAPAAASLGISTGDIVITA 306
Db 291 -----ELGIMGTELNSLAKAMKYDAQRGAFVSQVMNSNAKAGIKAGDIVITS 339
Qy 307 VDGAPINSATAMADALNGHHPGDVISVNMOTKSGGTRTGNVTAAE 351
Db 340 LNKGPISFAALRAQVGTMPVGSKISLG-LLREGKAITVNLLEQQ 383

RESULT 5
HTRA_LACHE
ID HTRA_LACHE STANDARD; PRT; 413 AA.
AC Q9Z4H7;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine protease do-like htra (EC 3.4.21.-).
GN HTRA.
OS Lactobacillus helveticus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1587;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=53/7;
RX MEDLINE=99047559; PubMed=9829922;
RA Smeds A., Varmanen P.K., Palva A.M.;
RT "Molecular characterization of a stress-inducible gene from
RT Lactobacillus helveticus."
RL J. Bacteriol. 180:6143-6153(1998).
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ005672; CAA06668.1; -
DR MEROPS; S01.273; -
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001940; Peptidase_SIC.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR PROSITE; PS0106; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
DR Hydrolase; Serine protease; Transmembrane.
FT DOMAIN 20 40
FT DOMAIN 141 302
FT DOMAIN 305 401
FT ACT SITE 150 150
FT ACT SITE 180 180
FT ACT SITE 258 258
FT ACT SITE 258 258
SQ SEQUENCE 413 AA; 42647 MW; B16B677991C88707 CRC64;
```

```
Query Match 17.8%; Score 321; DB 1; Length 413;
Best Local Similarity 31.3%; Pred. No. 3.2e-11;
Matches 89; Conservative 50; Mismatches 121; Indels 24; Gaps 8;
Qy 67 NTKLGNNAVAGAGTGIVI---DPNGVLTNNHVIAGATDINAFVSGSQTYGVDVVGYDR 123
Db 119 SSKNGKLETYSSEGGVVMKSKNGYIVTNNHVISGSDAVQVL-LANGKTVNAKVVGKDS 177
Qy 124 TODVAVLQIRAGGIPSAIAGG--VAVGEPVAMGNSGGGGTTPRAVGRVVALGQTVQ 181
Db 178 TTDLAIVSIDAKYVTTQAFGDSKHLAQGVIAVGSPLGSEYASTVTOGIIISAPARTIS 237
Qy 182 ASDSLTGAETLNGLIQFDAAIQPGDSGPPVNVGLQVVGWMT--AASDNFQSQGGQF 239
Db 238 TS---SGNQOT---VIQTDAAINFGNSGALVNSAGVIGNSMKLAOSSDGTSGVGNAP 291
Qy 240 AIPICQAMAIAGQIRSGG--GSPTVHIGPTAFLG-----LGVVDNNGNGARVORVVG 289
Db 292 AIPSNVETIVNELVKKGKITRPQLGVRVIALQGIPEGYSRLKIKSNLKNGIYIAFVSR 351
Qy 290 SAPAASLGISTGDIVITAVDGAIPINSATAMADALNGHHPGDVISV 333
Db 352 NGSANAGIKSGDIVITKVGKKVEDVASLHSLYKVGDTVNV 395

RESULT 6
DEGP_ECOLI
ID DEGP_ECOLI STANDARD; PRT; 474 AA.
AC P09376; P15724;
DT 01-MAR-1989 (Rel. 10, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protease do precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR PTD OR B0161 OR Z0173 OR ECS0165.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=89057448; PubMed=3057437;
RA Lipinska B., Sharma S., Georgopoulos C.;
RT "Sequence analysis and regulation of the htra gene of Escherichia
RT coli: a sigma 32-independent mechanism of heat-inducible
RT transcription."
RL Nucleic Acids Res. 16:10053-10067(1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=94261430; PubMed=8202364;
RA Fujita N., Mori H., Yura T., Ishihama A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 2.4-4.1 min (110,917-193,643 bp) region."
RL Nucleic Acids Res. 22:1637-1639(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA Davis K., Federspiel N., Hyman R., Kaiman S., Komp C., Kurdi O.,
RA Lashkari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA Davis R.W.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
```


123 KGVVTHNNVVDNATVIV-QLSDGRKFDKMWGKDRSDIALIQIONPKNLTAIKWADS 191
145 GGVAVGEPVVMGNSGGCGGTTPRVPGRVVALGQTVQASDLSLTGAE-ETLNGHLIQFDAAI 203
182 DALRVGDVTVAGNPFGLGET--VTSIGVSLGR-----SGLNAENYENFIQTDAAI 231
204 QPGDSGSPVNLGQVVGMMTA--ASDNFOLSGGQGFAP-----IGAMALAGQIRS 255
232 NRGNSGALVNLGELGINTAILAPDGGNI--GIGFAIPSNVKNLTSCMWVEYQGVKR 288
256 GGGSPVTHIGTAPLGLGVVDNNGN-----GARVQVVGSAFAASLGISTGD 302
289 G-----ELGIMGTELSLAKAMKVDAQRGAFVSVQVLPNSSAAKAGIKAGD 334
303 VITAVDGAIPNSATAM 318
335 VITSLNGKPISSFAAL 350

RESULT 7
HTRA_LACLA
ID HTRA_LACLA STANDARD; PRT; 408 AA.
AC Q9LA06;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Serine protease do-like htra (EC 3.4.21.-) (HtraL).
GN HTRA OR Lz136.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Manger S., Jaillon O., Malarne K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403";
RL Genome Res. 11:731-753(2001).
CC -!- FUNCTION: DEGRADES ABNORMAL EXPORTED PROTEINS. NEEDED FOR THE PRO-
CC PEPTIDE PROCESSING OF A NATURAL PRO-PROTEIN AND FOR MATURATION OF
CC A NATIVE PROTEIN. RESPONSIBLE FOR THE HOUSEKEEPING OF EXPORTED
CC PROTEINS.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Probable).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF155705; AAF61294.1; -;
CC EMBL; AE006442; AAK06234.1; -;
CC PIR; H86891; H86891.
CC MEROPS; S01.273; -;
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase S1.
CC InterPro; IPR001940; Peptidase_S1C.

DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PRO0089; trypsin; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Serine protease; Transmembrane; Complete proteome.
FT TRANSMEM 6 26
FT DOMAIN 88 284
FT ACT_SITE 127 127
FT ACT_SITE 157 157
FT ACT_SITE 239 239
SQ SEQUENCE 408 AA; 41648 MW; 581B90E5A7DF851 CRC64;
Query Match 17.2%; Score 310; DB 1; Length 408;
Best Local Similarity 31.3%; Pred. No. 1.3e-10;
Matches 93; Conservative 51; Mismatches 119; Indels 34; Gaps 11;
79 GTGIVDPNG---VVLTHNHIAGATDINAFVSGQYGVVDVGYDRTQDVAVLQLRG- 134
108 GSGVIYKSGDAYVYVTVNHYVHVIAGNSLDVLISG-GQVKASVVGVDYDVLAVLKISSE 166
135 -AGGLPSAAGGVAVGVPVAVGNSGGGCTTPRVPGRVVALGQTVQASDLSLTGAEETL 193
167 HVKDVATFADSSKLITGEPAIAVGSPLGSQFANTATEGILTSATSRQVLTQE-NGQTNI 225
194 NGLIQFDAAIQPGDSGSPVNLGQVVGMMTA--ASDNFOLSGGQGFAPIGQAMAIA 250
226 NA-IQTDRAINFNSGALINIEGQVIGITQSKITTTEDGTSVEGLGFAIPSNVDVNI 284
251 GQIRSGGSPVTHIGTAPLGLGVVD-----NNGN-----GARVQVVGSAFAA 294
285 NKLEADG-----KISRPA-LGIRMVLDLSQSLNDSQKLPSVTVGGVVVSVQGLPAA 338
295 SIGISGTDVITAVDGAIPNSATAMADALNGHHPGVDVSVNWQTKSGGTGTGNVTLAE 351
339 SAGLKAGDVITKVGDTAVTSSDLSALYSHNINTVKVTVY-RDGKSNATADVKLK 394

RESULT 8
DEGP BRUME STANDARD; PRT; 513 AA.
AC O8YG32;
DT 15-MAR-2004 (Rel. 43, Created)
DT 15-MAR-2004 (Rel. 43, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR EMEI1330.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / Biotype 1;
RX MEDLINE=20020109; PubMed=11756688;
RA DelVecchio V.G., Kapral V., Redkar R.J., Patra G., Mijer C., Los T.,
RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Renik G.,
RA Jablonski L., Larsen L., D'Souza M., Bernal A., Mazur M., Goldsman E.,
RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
RA Haselkorn R., Kyrides N., Overbeek R.;
RT "The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis";
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
CC -!- SUBCELLULAR LOCATION: Periplasmic (potential).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

ar 1;
K.E., Eisen J.A., Heidelberg J.F.,
Brinkac L.M., Beanan M.J.,
A.S., Kolonay J.F., Madupu R.,
Shetty J., Malek J.F., Van Aken S.
R., White O., Salzberg S.L.,
S.M., Boyle S.M., Fraser C.M.;
fundamental similarities between
bionts".
3148-13153(2002).
mic (Potential).
se family S2C.
domains.

. It is produced through a collaboration of informatics and the ENBL output. There are no restrictions as long as its content is in removed. Usage by and for content (See <http://www.isb-sib.ch/ab.ch>).

asmic; Repeat; Signal;

BLE SERINE PROTEASE DO-LIKE.
YTIC.

.
. E RELAY SYSTEM (POTENTIAL).
. E RELAY SYSTEM (POTENTIAL).
. E RELAY SYSTEM (POTENTIAL).
EF1959472806 CRC64;

305; DB 1; Length 513;
No. 3e-10;
smatches 142; Indels 136; Ga

WTAPA-----QAAPPALISQ
 |||||
 ALALAGAFVVTGPLGALNEAAVHVTPP--PQ-----
 -----TKKL-----
 |
 WKVKDVQETSNRGQFGFPGFDQIDGHPUKKR

ar 1;
K.E., Eisen J.A., Heidelberg J.F.,
Brinkac L.M., Beanan M.J.,
A.S., Kolonay J.F., Madupu R.,
Shetty J., Malek J.F., Van Aken S.
R., White O., Salzberg S.L.,
S.M., Boyle S.M., Fraser C.M.;
fundamental similarities between
bionts".
3148-13153(2002).
mic (Potential).
se family S2C.
domains.

. It is produced through a collaboration of informatics and the ENBL output. There are no restrictions as long as its content is in removed. Usage by and for content (See <http://www.isb-sib.ch/ab.ch>).

asmic; Repeat; Signal;

BLE SERINE PROTEASE DO-LIKE.
YTIC.

.
. E RELAY SYSTEM (POTENTIAL).
. E RELAY SYSTEM (POTENTIAL).
. E RELAY SYSTEM (POTENTIAL).
EF1959472806 CRC64;

305; DB 1; Length 513;
No. 3e-10;
smatches 142; Indels 136; Ga

WTAPA-----QAAPPALISQ
 |||||
 ALALAGAFVVTGPLGALNEAAVHVTPP--PQ-----
 -----TKKL-----
 |
 WKVKDVQETSNRGQFGFPGFDQIDGHPUKKR

QY 71 -----GNNVAGAGTGIVIDPVGVLNNHVIAGATDINAF 106
DB 104 DFGMEPRGSDNRKANKRPGREFVAGQSGFVSDGVYVNNHW---SDGDAY 160
QY 107 SV--GSGQTYGVVDVYDRTQDAVLQRLGAGLPS-----AAIG--GGVAVGEFVWANG 157
DB 161 TVVLDGTELDKALIGADPTDLAVLKINA-----PKRKEVYVAFGDDNKVYGVWVAVG 216
QY 158 NSGGCGGTTPRAGRVVAGQTVQASDSLTGAETLNGLIQDAIQAQDSGGPVVNGLG 217
DB 217 NPFGLGCT--VTSGIVSARGRDIGAG-----PYDDFIQIDAAYVKNKSGGPAFDLSG 266
QY 218 QVGVNTASDRAFDLSQGGG--FATPIGAMAIAGQIRSGG-----GSTVTHIGP----- 266
DB 267 EVIGINTAI---PSPGGVGVIAFAIPSTAKQVVDQLIKKSVRGWIGVQIOPVTKDI 323
QY 267 TAFGLGVVDNNGNGARVORVVGSAASLGHTGDTVITAVGAPINSATAMADALNGHH 326
DB 324 AASLGIA----BEKGAIVASPDQDGAAXAGIKAGDVITAVNGETVQDPRDLARKVANIA 379
QY 327 PGDVISVNWTKSGGRTGNTVLAEGP 353
DB 380 PGKAAITVVRKNKABEI-NVTIAAMP 405

RESULT 10
HTRA_BACSU STANDARD; PRT; 449 AA.
AC 034358;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable serine protease do-like htra (EC 3.4.21.-).
GN HTRA OR BSU12900.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Devine K.M.;
RT "Sequence of the Bacillus subtilis genome between xlyA and ykoR";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=38044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinis S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccori E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpsira P., Tognoni A.,
RA Toasato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,

RA Viari A., Wambutt R., Wedler H., Wedler E., Wedler K., Weitzneger T.,
RA Winters P., Wipat K., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis";
RL Nature 390:249-256(1997).
RN [3]
RP TRANSCRIPTIONAL REGULATION.
RC STRAIN=168;
RX MEDLINE=20158875; PubMed=10692364;
RA Noone D., Howell A., Devine K.M.;
RT "Expression of ykda, encoding a Bacillus subtilis homologue of Htra,
RT is heat shock inducible and negatively autoregulated";
RL J. Bacteriol. 182:1592-1599(2000).
RN [4]
RP TRANSCRIPTIONAL REGULATION.
RC STRAIN=168;
RX MEDLINE=20576168; PubMed=11133960;
RA Noone D., Howell A., Collier R., Devine K.M.;
RT "Ykda and ykda, Htra-like serine proteases in Bacillus subtilis,
RT engage in negative autoregulation and reciprocal cross-regulation of
RT ykda and ykda gene expression";
RL J. Bacteriol. 183:654-663(2001).
RN [5]
RP TRANSCRIPTIONAL REGULATION.
RC STRAIN=168;
RX MEDLINE=21439741; PubMed=1155295;
RA Hyttilainen H.-L., Bolhuis A., Darmon E., Muukkonen L., Koski P.,
RA Vitikainen M., Sarvas M., Pragai Z., Bron S., van Dijk J.M.,
RA Kontinen V.P.;
RT "A novel two-component regulatory system in Bacillus subtilis for the
RT survival of severe secretion stress";
RL Mol. Microbiol. 41:1159-1172(2001).
CC -!- FUNCTION: May be involved in processing, maturation, or secretion
CC of extracellular enzymes.
CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -!- INDUCTION: Transcription is cspS dependent. Induced by heat shock
CC during exponential growth and by heterologous amyloses at the
CC transition phase of the growth cycle. Negatively regulates its own
CC expression during exponential growth and during heat shock. Expression
CC of ykda, especially during stress conditions.
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ002571; CAA05570.1; -;
CC EMBL; Z99110; CAB13147.1; -;
CC MEROPS; S01.273; -;
CC Subtilisin; BG12608; htra.
CC InterPro; IPR009003; Cys_ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ; 1.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SM00228; PDZ; 1.
CC PROSITE; PS0106; PDZ; 1.
CC Hydrolyase; Protease; Serine protease; Heat shock; Transmembrane;
CC Complete proteome.
KW DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 45 67 POTENTIAL.
FT DOMAIN 68 449 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 103 108 POLY-SER.

Query Match 16.1%; Score 291; DB 1; Length 504;
Best Local Similarity 33.3%; Pred. No. 1.8e-09;

Matches 97; Conservative 39; Mismatches 117; Indels 38; Gaps 12;

QY 79 GTGIVIDPENGVLVLTNNHVIAGATDINAFSV--GGQTGYGVDPVVGVDRTQDVAVIQL--RG 134
.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.
Db 124 GSGFFITEDGYLVTTNHHVV---SDGSFTVMNDGTDLAKLVGKDSRTDLAVLKVDDKR 180

QY 135 ACGLPSAALGGVAIGEPFVWANGNSGGCGGTTPRAVPGRVVALGQTVQASDSLTAETLN 194
.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Db 181 KPTVVSFDADDEKRVGDVWVAVGNPFPGGGTVA--GIISARGRDIGSG-----PYD 230

QY 195 GLIQFDAAIQPDGSGPPVNGVLGGVGVGNMTAASNDFOLSQQGG--FAPIGQAAMIAQG 252
.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Db 231 DYLQYDAARNRNSGGPIFNLSGEVVGINTAI---FSPSGNVGIAFAIPASVAKDVVDS 287

QY 253 IRSGG-----GSTPVYHGFTA-----FLGLGVVDNNGARVQRVVGSPAASLGISTGDV 303
.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Db 288 LTKDGTSGVRMLGVQIQPVTKDIASLSGL-----SEANGALLVVEPAQSFGPKAGIKNGDV 343

QY 304 ITAVDGAPINSATAMADALNHHPHDVT-SVN-WQTKSGGTTCNTVTLAEGP 353
.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.: :.:
Db 344 VTALNGEVPKDRLARRVAALRPSTAEVTLM--RSKGSETVNLIEIGTLP 392

RESULT 13

DEGP BARHE STANDARD; PRT; 503 AA.

AC P54925,
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable periplasmic serine protease DO-like precursor (EC 3.4.21.-) (Antigen htra).

DN DEGP OR HTRA.
OS Bartonella henselae (Rochalimaea henselae).
OC Bacterii; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bartonellaceae; Bartonella.
ON NCBI_TaxID=38323;
RX [1]

RN SEQUENCE FROM N.A.
RP STRAIN=ATCC 49882 / Houston 1;
RC MEDLINE=94299828; PubMed=8027347;
RX Anderson B., Sims K., Regnery R., Robinson L., Schmidt M.J.,
RA Gorai S., Rager C., Edwards K.;
RT "Detection of Rochalimaea henselae DNA in specimens from cat scratch disease patients by PCR."
RN J. Clin. Microbiol. 32:942-948(1994).
RL CC -/- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -/- SIMILARITY: Belongs to peptidase family S2C.
CC -/- SIMILARITY: Contains 2 PDZ/DHR domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

CC -----
DR EMBL; L20127; AAA97430.1; -.
DR MEROPS; S01.273; -.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ_1.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_SIC.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
KW Hydrolase; Serine protease; Periplasmic; Repeat; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 503 PROBABLE PERIPLASMIC SERINE PROTEASE DO-LIKE.

FT DOMAIN 286 357 PDZ 1.
 FT DOMAIN 419 466 PDZ 2.
 FT ACT SITE 143 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 173 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT SITE 247 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 503 AA; 54114 MW; 6CD9F473282AF9E CRC64;

Query Match
 Best Local Similarity 16.1%; Score 290; DB 1; Length 503;
 Matches 113; Conservative 42; Mismatches 135; Indels 118; Gaps 16;

QY 22 GLGLATAPAAQ---APPALSDQRFADFPALPLDPSAMVAQVAVQVNTKLG----- 71
 DB 29 GSSLWTKAHANSVFSSLMQQQGFAD-----IVSQKPAVVSQVSKNKKKEWF 78
 QY 72 -----YN-----NAVGAGTGIVDPNGV 90
 DB 79 FSDPSTPGFDLPQHPKXFFQFYARDKPSNKSLSORSHRLPIAPSGFFISSDGI 138
 QY 91 LNNHVIAGATDINAFSVSGQTYGVVDYDRTQDVAVLQRLGAGGLPSAAIG--GGVA 148
 DB 139 VTNHVISDGTSY-AVLLDDGTGLNKLIGDPTDLAVLKVKNEKRFYSYVDFGDSKLR 197
 QY 149 VGEPIVAMNSGGQGTFRVGRVAVGQVQASDLSLTGAETLNGLIQDAAIQGDS 208
 DB 198 VGDWVVALGNPFLGGTVA--GIVSARGR-----DIGTG---VYDFFIQIDAAVNRGNS 247
 QY 209 GPPVNVGLGVVGMNTAASDNFQSQGGQ--FAIPIGQAMAIAGAQIRSGGSGPTVHGP 266
 DB 248 CGPTFDLNGKVVGVNTAI--FSPSGNVGIAFAIPANTANEVVQQL----- 291
 QY 267 TAFGLGLGVNDNGCARVQV-----VG-----SAPASLGISITGDVITA 306
 DB 292 ---IEKGLVQRGWLGVIQPTKIEISIGLKEAKGALITDPLKGPAAKAGIKAGDVIIS 348
 QY 307 VDGAPINSATAMADALNGHPGDVIVSN-WQTKSGGTRTGNTVLAEGP 353
 DB 349 VNGEKINDVRDLAKRIANWSFGETVTLGW--KSGKSENINKVLDSP 394

RESULT 14

ID DEGI ARATH STANDARD; PRT; 437 AA.
 AC O22609; Q9LKB5.
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Protease Do-like 1, chloroplast precursor (EC 3.4.21.-).
 GN DEGP1 OR DEGP OR AT3G27925 OR K16N12.18.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 ON NCBI_TaxID=3702;
 RX MEDLINE=98175982; PubMed=9507020;
 RA Itzhaki H., Naveh L., Lindahl M., Cook M., Adam Z.;
 RT "Identification and characterization of DegP, a serine protease
 associated with the luminal side of the thylakoid membrane.";
 RN J. Biol. Chem. 273:7094-7098(1998).
 RN [2]
 RC SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20363099; PubMed=10907853;
 RA Kaneko T., Katoh T., Sato S., Nakamura A., Asamizu E., Tabata S.;
 RT "Structural analysis of Arabidopsis thaliana chromosome 3. II.
 RT Sequence features of the 4,251,695 bp regions covered by 90 P1, TAC
 RL and BAC clones.";
 RN DNA Res. 7:217-221(2000).
 RN [3]
 RP SEQUENCE OF 104-118.
 RC STRAIN=cv. Columbia;

RA Kieselbach T., Bystedt M., Schroeder W.P.;
 RL Submitted (JUL-2000) to Swiss-Prot.
 CC -!- FUNCTION: Serine protease that is required at high temperature.
 CC May be involved in the degradation of damaged proteins. In vivo,
 CC can degrade beta-casein.
 CC -!- ENZYME REGULATION: Inhibited by phenylmethylsulfonyl fluoride and
 CC O-phenanthroline.
 CC -!- SUBCELLULAR LOCATION: BOUND TO LUMINAL SIDE OF THE THYLAKOID
 CC MEMBRANE.
 CC -!- INDUCTION: By heat shock.
 CC -!- SIMILARITY: Belongs to peptidase family S2C.
 CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF028842; AAC39436.1; --
 CC EMBL; AF000371; BAB02539.1; --
 CC EMBL; AF001302; BAB02539.1; JOINED.
 CC MEROPS; S01.279; --
 CC InterPro; IPR009003; Cys_Ser_trypsin.
 CC InterPro; IPR001478; PDZ.
 CC InterPro; IPR001254; Peptidase_S1.
 CC InterPro; IPR001940; Peptidase_S1C.
 CC Pfam; PF00595; PDZ; 1.
 CC Pfam; PF00889; trypsin; 1.
 CC PRINTS; PRO0834; PROPEASES2C.
 CC SMART; SM00228; PDZ; 1.
 CC PROSITE; PS0106; PDZ; 1.
 KW Hydrolyase; Serine protease; Transit peptide; Chloroplast; Thylakoid.
 FT TRANSIT 1 ?
 FT CHAIN 104 437
 FT DOMAIN 152 321
 FT DOMAIN 324 421
 FT ACT SITE 171 171
 FT ACT SITE 201 201
 FT ACT SITE 280 280
 FT CONFLICT 12 23
 FT CONFLICT 36 36
 FT CONFLICT 54 54
 FT CONFLICT 60 60
 FT CONFLICT 64 64
 FT CONFLICT 68 69
 FT CONFLICT 355 355
 FT CONFLICT 381 381
 FT CONFLICT 416 416
 SQ SEQUENCE 437 AA; 46213 MW; 1497BIAB3F5FF2A4 CRC64;
 Query Match
 Best Local Similarity 16.1%; Score 289.5; DB 1; Length 437;
 Matches 111; Conservative 56; Mismatches 143; Indels 81; Gaps 16;

QY 7 RSLRWSLLSLVLA-----VGLGLATAPAAQAAPP-----ALSQDRFADPFA----- 47
 DB 42 RSKRYFRILSKPLNDNNGDDGDTLLTTPESAVKPFLLCTSVALSFSLFAAPSAVES 101
 QY 48 -----LPLDPSAMV---AQVAPQVNTKLGNNAV-----GAGTGVID 85
 DB 102 SAFVSTPKKLOTDELATVRFQENTSVVITNLAVRQDAFTLDVLEVPQSGSGFW 161
 QY 86 PNGVLTNNHVIAGATDINAFSVSGQTYGVVDYDRTQDVAVLQRLGAGG----LP 141
 DB 162 KQHIVTNYHVIRGASDLRV--TLADQTTFDKAVGVGFDQKDAVLRIDAPKNLRIP-V 219
 QY 142 AIGGGVAVGEPVAMNSGGQGTFRVGRVAVGQVQASDLSLTGAETLNGLIQF 201

Db 220 GVSADLLVCQKVPATGNPFLDHT--LITGVISGLRREI--SSAATG--RPIQDVQIOTDA 273
QY 202 AIQPGDSGGVYVNGVGMNTAASDNFQSGGQGFPAIPGQAWAIAQIRSGGSGPT 261
Db 274 AINPGNSGFLDSSGTLIGINTAIYSPGAS--SGVGFSPVDVTGIVDQL----- 324
QY 262 VNIQPTAFGLGV-----VDNNG--NGARVQVVGSAAPASLIGST-----GD 302
Db 325 VFGKVTIRILGKFPADQSVQQLGVSGVLLDAPPSGPGAKGLQSTKRDGYGRLLIGD 384
QY 303 VITAVDGPAPINSATAMADALNGHHPGDVLSV 333
Db 385 IITSVNGTKVNSGDIYRLDCKQGVDEVTV 415

RESULT 15

DEGP CHLTR STANDARD; PRT; 497 AA.
ID AC PI8584; 084830;
DT 01-NOV-1990 (Rel. 16, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-) (59 kDa
DE immunogenic protein) (SK59).
GN DRGP OR HTRA OR CT823.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=Serovar L2;
RC MEDLINE=90337348; PubMed=2379836;
RX Kahane S., Weinstein Y., Sarov I.;
RA "Cloning, characterization and sequence of a novel 59-kDa protein of
RT Chlamydia trachomatis."
RL Gene 90:61-67(1990).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=D/UW-3/CX;
RC MEDLINE=9900809; PubMed=9784136;
RX Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis."
RL Science 282:754-759(1998).
CC -1- SIMILARITY: Belongs to peptidase family S2C.
CC -1- SIMILARITY: Contains 2 PDZ/DHR domains.
CC -1- CAUTION: REF.1 SEQUENCE WAS INCORRECT: DUE TO SEQUENCING ERRORS
CC THE AUTHORS TRANSLATED THEIR PUTATIVE 59 kDa IMMUNOGENIC PROTEIN
CC ON THE COMPLEMENTARY STRAND TO THAT OF WHAT SEEMS TO BE THE REAL
CC PROTEIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AE001355; AAC68420.1; -.
CC EMBL; M31119; AAA23116.1; -.
CC PIR; H71465; H71465.
CC PHCI-2DPAGE; P18584; -.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ-
CC InterPro; IPR001254; Peptidase S1.
CC InterPro; IPR008256; Peptidase_S1B_v8.
CC InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ; 2.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00834; PROTEASES2C.

PRINTS; PR00839; V8PROTEASE.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS0106; PDZ; 2.
KW Hydrolase; Serine protease; Repeat; Signal; Antigen;
KW Complete proteome.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 497 PROBABLE SERINE PROTEASE DO-LIKE.
FT DOMAIN 128 289 CATALYTIC.
FT DOMAIN 290 381 PDZ 1.
FT DOMAIN 394 485 PDZ 2.
FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 497 AA; 53244 MW; 86A5E31BB84A38BA CRC64;
Query Match 16.0%; Score 288; DB 1; Length 497;
Best Local Similarity 32.3%; Pred. No. 2.6e-09;
Matches 95; Conservative 46; Mismatches 109; Indels 44; Gaps 12;
QY 79 GTGIVDPNGVLTNNHVIAGTDINAFSGSGGTGVVGYDRTQDVAVLQRGAGGL 138
Db 127 GTGIVSEGGVVTNNHVVEDAGKIHV--TLHDGOKYAKIVGLDPKTDLAVIKIQ--AEKL 184
QY 139 PSAAIGGG--VAGCEPVVAMNSGGGQGTTPRAVGRVVALGOT---VQASDLSLTGAET 192
Db 185 PFLTSGNSDQLQIGDWAIAIGNPFLQAT--VTGVVISAKGRNQLHIVDFED----- 234
QY 193 LNGLIQFDAIQTQDGGSGPVVNGLVGVGMNTAASDNFQLSQG--GGGFAIPICQAMAIA 250
Db 235 ---FIQTDALINPGNSGGPLININGVIGVNTAIVSG---SGGYIGIGFAIPSLMAKRV 288
QY 251 QGIRSGGSPVTHIGTPAFLGLGVVDN-----NNGARVQVVGSAAPASLIGSTG 301
Db 289 DQLISDGGQVTRGFLGVT---LQPIDSELATCYKLEKYGVGALVTDVVGKSPAELKGRQE 344
QY 302 DVITAVDGPAPINSATAMADALNGHHPGD--VISVNWOTKSGGTRTGNVTLAEGP 353
Db 345 DVIVAYNGKEVESLSALRNAISLMMPGTRVVKI---VREGKTIETPVTVTQIP 395

Search completed: June 30, 2004, 16:49:41
Job time : 11.7798 secs

1. The first part of the document is a letter from the President of the United States to the Congress, dated January 1, 1801. It is a very important document, as it is the first time that the President has addressed the Congress since the establishment of the new government. The letter is written in a very formal and dignified style, and it contains many important points. The President begins by expressing his gratitude to the Congress for the honor of electing him to the office of President. He then goes on to discuss the state of the Union, and the progress of the new government. He mentions the many difficulties that have been overcome, and the many successes that have been achieved. He also discusses the future of the country, and the steps that he plans to take to ensure the prosperity and happiness of the people. The letter is a very important document, as it sets the tone for the new government, and it outlines the President's vision for the future of the country.

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

GenCore version 5.1.6
Copyright (C) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:01 ; Search time 25.8569 Seconds
(without alignments)
4331.879 Million cell updates/sec

Title: US-09-597-796C-4

Perfect score: 1802

Sequence: 1 MSNRRRLRWLWLLVLA.....QTKSGGTRGNVTLAEGPPA 355

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mnc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1792	99.4	355	007175	007175 mycobacteri
2	1792	99.4	355	16 Q7U2S9	Q7U2S9 mycobacteri
3	1306	72.5	361	2 Q50320	Q50320 mycobacteri
4	1272.5	70.6	354	16 Q9CCY9	Q9CCY9 mycobacteri
5	469.5	26.1	445	16 Q8VKA4	Q8VKA4 mycobacteri
6	469.5	26.1	464	16 Q7U0X2	Q7U0X2 mycobacteri
7	468.5	26.0	464	16 Q53896	Q53896 mycobacteri
8	460	25.5	382	16 Q9CD67	Q9CD67 mycobacteri
9	460	25.5	452	2 Q925G6	Q925G6 mycobacteri
10	407	22.6	542	16 Q9FBK9	Q9FBK9 streptomyc
11	396	22.0	519	16 Q93J30	Q93J30 streptomyc
12	391	21.7	375	16 Q8DG87	Q8DG87 synchococc
13	382	21.2	394	16 P72780	P72780 synchocyst
14	382	21.2	472	16 Q82FM9	Q82FM9 streptomyc
15	382	21.2	619	16 Q82IL8	Q82IL8 streptomyc
16	379.5	21.1	473	16 Q8PR17	Q8PR17 corynebacte

17	373.5	20.7	500	2 Q9KJN6	Q9KJN6 myxococcus
18	372	20.6	407	16 Q8YTF9	Q8YTF9 anabaena sp
19	366	20.3	525	16 Q89QJ8	Q89QJ8 bradyrhizob
20	364.5	20.2	675	16 Q8G6T3	Q8G6T3 bifidobacte
21	363	20.1	395	16 Q7V5C8	Q7V5C8 prochloroco
22	362	20.1	525	16 Q8PMV4	Q8PMV4 xanthomonas
23	358	19.9	413	16 Q8NS10	Q8NS10 corynebacte
24	355.5	19.7	452	16 P73354	P73354 synchocyst
25	354.5	19.7	490	16 Q8XPT5	Q8XPT5 ralsstonia s
26	354	19.6	371	16 Q31388	Q31388 bradyrhizob
27	354	19.6	433	16 Q97GD5	Q97GD5 clostridium
28	353	19.6	325	16 Q8DMV9	Q8DMV9 synchococc
29	349	19.4	447	16 Q8R756	Q8R756 thermoaer
30	347.5	19.3	391	16 Q895I5	Q895I5 clostridium
31	347.5	19.3	432	16 Q7U8K9	Q7U8K9 synchococc
32	347	19.3	374	16 Q7U495	Q7U495 synchococc
33	347	19.3	465	16 Q92QE6	Q92QE6 rhizobium m
34	346	19.2	525	16 Q8PB56	Q8PB56 xanthomonas
35	346	19.2	389	16 Q8DL28	Q8DL28 synchococc
36	345	19.1	339	16 Q89RP2	Q89RP2 bradyrhizob
37	345	19.1	362	16 Q9S2K5	Q9S2K5 streptomyc
38	344	19.1	511	2 Q8RTK2	Q8RTK2 xanthomonas
39	343	19.0	457	16 Q8ZB58	Q8ZB58 yersinia pe
40	343	19.0	463	16 Q8DIR0	Q8DIR0 yersinia pe
41	342.5	19.0	528	16 Q89321	Q89321 bradyrhizob
42	341.5	19.0	514	16 Q9PBA3	Q9PBA3 xylella fas
43	340.5	18.9	505	16 Q8Y0I6	Q8Y0I6 ralsstonia s
44	338.5	18.8	629	16 Q7U0X4	Q7U0X4 rhodopirell
45	335.5	18.6	429	16 Q8YYZ0	Q8YYZ0 anabaena sp

ALIGNMENTS

RESULT 1

007175	PRELIMINARY;	PRT;	355 AA.
ID	007175	007175	
AC	007175:		
DT	01-JUL-1997 (TREMELrel. 04, Created)		
DT	01-JUL-1997 (TREMELrel. 04, Last sequence update)		
DT	01-OCT-2003 (TREMELrel. 25, Last annotation update)		
DE	Hypothetical protein (Serine protease, putative).		
GN	PEPA OR RV0125 OR MTCI418B.07 OR MT0133.		
OS	Mycobacterium tuberculosis.		
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;		
OC	Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.		
OX	NCBI_TaxID=1773;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=H37RV;		
RX	MEDLINE=98295987; PubMed=9634230;		
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,		
RA	Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,		
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,		
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,		
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,		
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,		
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,		
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;		
RA	"Complete genome sequence."		
RT	Nature 393:537-544(1998).		
RL	[2]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=CDC 1551 / Oshkosh;		
RC	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,		
RA	Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,		
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,		
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,		
RA	Bishai W.;		
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and		
RT	laboratory strains."		
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.		

CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL; Z96071; CAB09453.1; -.
 DR EMBL; AE006925; AAK44357.1; -.
 DR PIR; F70983; F70983.
 DR TIGR; MT0133; -.
 DR TubercuList; Rv0125; -.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR008256; Peptidase_S1B V8.
 DR InterPro; IPR001940; Peptidase_SIC.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR PRINTS; PR00839; V8PROTEASE.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydroxylase; Hypothetical protein; Serine protease; Protease;
 KW Complete proteome.
 SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match 99.4%; Score 1792; DB 16; Length 355;
 Best Local Similarity 99.4%; Pred. No. 3.5e-87;
 Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MNSRRSLRWSSWLLSLVLAAGVGLATAPAAQAPFALPDSPAMVAQVA 60
 DB 1 MNSRRSLRWSSWLLSLVLAAGVGLATAPAAQAPFALPDSPAMVAQV 60
 QY 61 PQQVNTKLGYNNAVAGTGVIDPQVLTNNHVIAGATDINAFSGSGQTYGVDVVG 120
 DB 61 PQQVNTKLGYNNAVAGTGVIDPQVLTNNHVIAGATDINAFSGSGQTYGVDVVG 120
 QY 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTFRVAVGRVVALGQTV 180
 DB 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTFRVAVGRVVALGQTV 180
 QY 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGNTAASDNFOLSGGGGPA 240
 DB 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGNTAASDNFOLSGGGGPA 240
 QY 241 IPIQAMAIAQIRSGGSPVTHIGPTAFGLGVVDNNGGARVQVVGAPAAASLGIST 300
 DB 241 IPIQAMAIAQIRSGGSPVTHIGPTAFGLGVVDNNGGARVQVVGAPAAASLGIST 300
 QY 301 GDVITAVDGAPINSATAMADALNGHPGDVSVNMQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGAPINSATAMADALNGHPGDVSVNMQTKSGGTRTGNVTLAEGPPA 355

RESULT 2

QY0259 PRELIMINARY; PRT; 355 AA.
 ID QY0259
 AC QY0259;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE Probable serine protease PEPA (EC 3.4.21.-).
 GN PEPA OR W0130.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 ON NCBI_TaxID=1765;
 RX STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Biglieri K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,

RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RT "The complete genome sequence of Mycobacterium bovis.",
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL; BX248334; CAD92991.1; -.
 DR Hydroxylase; Complete proteome.
 SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match 99.4%; Score 1792; DB 16; Length 355;
 Best Local Similarity 99.4%; Pred. No. 3.5e-87;
 Matches 353; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MNSRRSLRWSSWLLSLVLAAGVGLATAPAAQAPFALPDSPAMVAQVA 60
 DB 1 MNSRRSLRWSSWLLSLVLAAGVGLATAPAAQAPFALPDSPAMVAQV 60
 QY 61 PQQVNTKLGYNNAVAGTGVIDPQVLTNNHVIAGATDINAFSGSGQTYGVDVVG 120
 DB 61 PQQVNTKLGYNNAVAGTGVIDPQVLTNNHVIAGATDINAFSGSGQTYGVDVVG 120
 QY 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTFRVAVGRVVALGQTV 180
 DB 121 YDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTFRVAVGRVVALGQTV 180
 QY 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGNTAASDNFOLSGGGGPA 240
 DB 181 QASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGNTAASDNFOLSGGGGPA 240
 QY 241 IPIQAMAIAQIRSGGSPVTHIGPTAFGLGVVDNNGGARVQVVGAPAAASLGIST 300
 DB 241 IPIQAMAIAQIRSGGSPVTHIGPTAFGLGVVDNNGGARVQVVGAPAAASLGIST 300
 QY 301 GDVITAVDGAPINSATAMADALNGHPGDVSVNMQTKSGGTRTGNVTLAEGPPA 355
 DB 301 GDVITAVDGAPINSATAMADALNGHPGDVSVNMQTKSGGTRTGNVTLAEGPPA 355

RESULT 3

QY0320 PRELIMINARY; PRT; 361 AA.
 ID QY0320
 AC QY0320;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DE 34kDa protein precursor.
 DE 34kDa protein precursor.
 OS Mycobacterium paratuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 ON NCBI_TaxID=1770;
 RX SEQUENCE FROM N.A.
 RC STRAIN=JD88/107;
 RC MEDLINE=95005449; PubMed=7921248;
 RX Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.,
 RT "Identification and characterisation of a putative serine protease
 expressed in vivo by Mycobacterium avium subsp paratuberculosis.",
 RL Microbiology 140:1977-1982(1994).
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
 DR EMBL; Z23092; CAA80638.1; -.
 DR PIR; S47170; S47170.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001940; Peptidase_SIC.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.

KW Hydrolase; Protease; Serine protease; Signal.
FT SIGNAL 1 38
SQ SEQUENCE 361 AA; 35709 MW; 30FEF78FD6F3C411 CRC64;
Query Match 72.5%; Score 1306; DB 2; Length 361;
Best Local Similarity 70.4%; Pred. No. 1.6e-61;
Matches 255; Conservative 41; Mismatches 58; Indels 8; Gaps 3;
QY 1 MSNS-RRSLRWLSLVLAAVCLGH-----ATAPAAAPPALSDQDFADFPALPLDPS 53
DB 1 MSKSHHRSVWSWLVGVLTVVGLGSLGSLGVLAPASAPSGALDRFADRLPLDPS 60
QY 54 AMVAQVAPQVNNITKLYNNNAVAGCTGIVDPNGVLTNNHVIAGATDINAFSVGSGQT 113
DB 61 AMVGQVGPQVNNITKLYNNNAVAGCTGIVDPNGVLTNNHVIAGATDINAFSVGSGQT 120
QY 114 YGVVVYGVYDRTQVAVLQRLGAGLPSAAGGVAVGEPVWAMGNSGGGGTTPRAVPGRV 173
DB 121 YAVDVYGVYDRTQVAVLQRLGAGLPSAAGGVAVGEPVWAMGNSGGGGTTPRAVPGRV 180
QY 174 VALGQTVQASDLSLTGAETLNGLIQDAATQPGDSGGPVVNGLGQVVGNTAASDNFOLS 233
DB 181 VALNQSVAIDTLTGAQENLGLLIQADAPIKFGDSGGPVWNSAGQVIGVDTAATDSYKMS 240
QY 234 QGGGFAIPIGQAMAIAGQIRSGGGGPTVHIGTAFILGLGVVDNNGNGARVQVWCSAFA 293
DB 241 -GGGFAIPIGRAMAVANQIRSGAGSNVTHIGTAFILGLGVVDNNGNGARVQVWNTGPA 299
QY 294 ASLGISGTVITAVDGPAPINSATADALNGHHPGVDIVSNWOTKSGGTRTGNVTLAEGP 353
DB 300 AAGIAPGDIVTGVDTVPINGAISMTVELVPHHPGVTIAVFRSVDGGERTANITLAEGP 359
QY 354 PA 355
DB 360 PA 361
RESULT 4
Q9CCY9 ID Q9CCY9 PRELIMINARY; PRT; 354 AA.
AC Q9CCY9
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable secreted serine protease.
GN ML2659.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL583926; CAC32191.1; -.
DR PIR; A87242; A87242.
DR Leproma; ML2659; -.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.

DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR008256; Peptidase S1B V8.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00083; trypsin; 1.
DR PRINTS; PRO0834; PROTEASES2C.
DR PRINTS; PRO0839; V8PROTEASE.
DR SMART; SMO228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 354 AA; 35265 MW; 612F23261BC9E4A CRC64;
Query Match 70.6%; Score 1272.5; DB 16; Length 354;
Best Local Similarity 70.1%; Pred. No. 9e-60;
Matches 251; Conservative 42; Mismatches 58; Indels 7; Gaps 3;
QY 1 MSNSRRSLRWLSLVLAAVGLGLATAQAQ---APPALSDQDFADFPALPLDPSAMVA 57
DB 1 MSRQHRSLRWLSLVLAAVGLGLATAQAQ---APPALSDQDFADFPALPLDPSAMVA-- 58
QY 58 QVAPQVNNITKLYNNNAVAGCTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTGYVD 117
DB 59 -VAPQVNNITKLYNNNAVAGCTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTGYVD 117
QY 118 VGVYDRTQVAVLQRLGAGLPSAAGGVAVGEPVWAMGNSGGGGTTPRAVPGRVVALG 177
DB 118 VGVYDRTQVAVLQRLGAGLPSAAGGVAVGEPVWAMGNSGGGGTTPRAVPGRVVALG 177
QY 178 QTVQASDLSLTGAETLNGLIQDAATQPGDSGGPVVNGLGQVVGNTAASDNFOLSQGG 237
DB 178 QTVQASEPLTGAETLNGLIQDAATQPGDSGGPVVNSRGQVVGNTAATDNYKN-LGGQ 236
QY 238 GFAIPIGQAMAIAGQIRSGGGGPTVHIGTAFILGLGVVDNNGNGARVQVWCSAFA 297
DB 237 GFAIPIGQAMAIAGQIRSGAGSNVTHIGTAFILGLGVVDNNGNGARVQVWNTGPA 296
QY 298 ISTGDIVITAVDGPAPINSATADALNGHHPGVDIVSNWOTKSGGTRTGNVTLAEGPPA 355
DB 297 ISVGLIITSVDGPPISEATATNVLVPHHPGVTIAVFRSVDGGDLTANVTLAEGPPA 354
RESULT 5
Q8VKA4 ID Q8VKA4 PRELIMINARY; PRT; 446 AA.
AC Q8VKA4
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Heat shock protein HtrA, putative.
GN MT1011.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gail J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE006985; AA045259.1; -.
DR TIGR; MT1011; -.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.

DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 DR SEQUENCE 446 AA; 44484 MW; 54170CBEA8FE872B CRC64;
 Query Match 26.1%; Score 469.5; DB 16; Length 446;
 Best Local Similarity 36.0%; Pred. No. 2.9e-17;
 Matches 127; Conservative 59; Mismatches 136; Indels 31; Gaps 10;
 QY 19 AAVGLGLATAPQAAPPALSDQADFPALPLDPSA---MVAQVAPQVNVNITKLGYNNA 75
 DB 102 AASLVGFNRPAPGSGGPPVAASAAPSIPAANNPPGSGVEQAAKVPFVVMLETLGRQSE 161
 QY 76 VGAGTGIVDPNGVLTNNHVIAGAT-----DINAFSGSGQTYGVVDVVGVDRTQDV 127
 DB 162 --EGSGIILSAEGLILTNHVIHIAAAKAPLGGPPPKTTVTFSDGRTAPTFTVVGADPTSDI 219
 QY 128 AVLQIRGAGLPSAAIGGG--VAVGPPVAMGNSGGGTPRAVGRVVALGQTVQASDS 185
 DB 220 AVRVQGVSGGLTPIISLGSSDLRVGPVLAIGSPGLGEGT--VTTGIVSALNRPVSTTGE 277
 QY 186 LTGAETLNGLIQDAAIOPGDSGGPVNGLGOVGMNTA-----ASDNFQLSQG--GGGF 239
 DB 278 --AGNQNTVLDAITDRAINPGNSGGALVNMNQLVGNNSAIATLGADSDAQSGLGLGF 336
 QY 240 AIPQQAIAAGIIRSGGSPVTHIGPTAFGLGVV--DNNNGGARVQVRVGSAPASLGI 298
 DB 337 AIPVQAKRIADELISTGKA-----SHASLGQVQVINDKXDPGAKIVEVAGGAANAGV 390
 QY 299 STGDVITAVDGPINSATAMADALNHHPGDVIVSNWTKSGGTRTGNVTIAE 351
 DB 391 PKGVVTVKDDRPINSADALVAARSKAPGATVALTFFQDPGSGSRTVQVTLGK 443
 RESULT 6
 QYUOX2 PRELIMINARY; PRT; 464 AA.
 AC QYUOX2
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Probable serine protease (Serine proteinase) (EC 3.4.21.-).
 GN MB1009.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OC NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97; PubMed=12788972;
 RA Garner T., Bigmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RA "The complete genome sequence of Mycobacterium bovis";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
 DR EMBL; BX248337; CAD93870.1; -.
 DR EMBL; Complete proteome.
 KW Hydrolyase; Complete proteome.
 SQ SEQUENCE 464 AA; 46436 MW; AE93A4BB3FFA9BE3 CRC64;
 Query Match 26.1%; Score 469.5; DB 16; Length 464;
 Best Local Similarity 36.0%; Pred. No. 2.9e-17;
 Matches 127; Conservative 59; Mismatches 136; Indels 31; Gaps 10;
 QY 19 AAVGLGLATAPQAAPPALSDQADFPALPLDPSA---MVAQVAPQVNVNITKLGYNNA 75
 DB 120 AASLVGFNRPAPGSGGPPVAASAAPSIPAANNPPGSGVEQAAKVPFVVMLETLGRQSE 179

QY 76 VGAGTGIVDPNGVLTNNHVIAGAT-----DINAFSGSGQTYGVVDVVGVDRTQDV 127
 DB 180 --EGSGIILSAEGLILTNHVIHIAAAKAPLGGPPPKTTVTFSDGRTAPTFTVVGADPTSDI 237
 QY 128 AVLQIRGAGLPSAAIGGG--VAVGPPVAMGNSGGGTPRAVGRVVALGQTVQASDS 185
 DB 238 AVRVQGVSGGLTPIISLGSSDLRVGPVLAIGSPGLGEGT--VTTGIVSALNRPVSTTGE 295
 QY 186 LTGAETLNGLIQDAAIOPGDSGGPVNGLGOVGMNTA-----ASDNFQLSQG--GGGF 239
 DB 296 --AGNQNTVLDAITDRAINPGNSGGALVNMNQLVGNNSAIATLGADSDAQSGLGLGF 354
 QY 240 AIPQQAIAAGIIRSGGSPVTHIGPTAFGLGVV--DNNNGGARVQVRVGSAPASLGI 298
 DB 355 AIPVQAKRIADELISTGKA-----SHASLGQVQVINDKXDPGAKIVEVAGGAANAGV 408
 QY 299 STGDVITAVDGPINSATAMADALNHHPGDVIVSNWTKSGGTRTGNVTIAE 351
 DB 409 PKGVVTVKDDRPINSADALVAARSKAPGATVALTFFQDPGSGSRTVQVTLGK 461
 RESULT 7
 QYUOX2 PRELIMINARY; PRT; 464 AA.
 AC QYUOX2
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Putative serine protease.
 GN RV0983 OR MTV044.11.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OC NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RC MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL; AL021999; CAA17582.1; -.
 DR PIR; C70821; C70821.
 DR Tuberculinist; RV0983;
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ_Ser_trypsin.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 KW Hydrolyase; Protease; Serine protease; Complete proteome.
 SQ SEQUENCE 464 AA; 46452 MW; AE93BFCC53E1EC8F CRC64;

Query Match 26.0%; Score 468.5; DB 16; Length 464;
 Best Local Similarity 36.0%; Pred. No. 3.3e-17;
 Matches 127; Conservative 59; Mismatches 136; Indels 31; Gaps 10;

QY 19 AAVGLGLATAPAAQAAPPALSDQFADFPALPLDPSA---MVAQVAPQVWVNTKLGYNNA 75
Db 120 AASLGFENRPAAGPGSGFVAASAAPSIPAANWPGSVQVAAKVVPVVMLETDLGEQSE 179
QY 76 VQAGTGVIDPQVNVLTNNHVIAGT-----DINAFSVGSGQTVGVVDVDRQDV 127
Db 180 --EGSGIILSAEGLITNNHVIAAAKPPLGSPPPKTTVTFSDDGRTAPFTVWGADPTSDI 237
QY 128 AVLQRLGAGLPSAIAIGGG--VAVGEPVVMWNGSGGQGTFRVPGRVWALGQTVQASDS 185
Db 238 AVRVQVSGVGLTPISLGSDDLVRGQVLAIGSPGLGEGT--VTTGVVSALNRVSVITGE 295
QY 186 LFGAETLNGLIQFPAALOPDGSQGVVNGLGQVVGWNTA-----ASNFLQSQ--GQGF 239
Db 296 -AGNQNTVLDAIQTDAAINPGNSGALVNMNAQLVGVNSAIALTGADSDAQSGSIGLGF 354
QY 240 AIPIGQMAIAQIRSGGSPVHIGTAFILGLGV--DNNNGARVORVVGSAAPASLGI 298
Db 355 ALFVQAKRIADLSTGKA-----SHASLGQVQVNDKDLGAKIVEVWAGNAAGV 408
QY 299 STGDVITAVDGPAPINSATAMADALNGHHPGDVSVNMQTKSGGTRTGNVTLAE 351
Db 409 FXGVVVVTKVDDRPINSADALVAARSKAPGATVALTFQDPGSGSRTVQVTLGK 461

RESULT 8
Q9CD67 PRELIMINARY; PRT; 382 AA.
ID Q9CD67
AC Q9CD67
DT 01-JUN-2001 (TREMELrel. 17, Created)
DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Possible secreted serine protease.
GN ML0176.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holtzoy S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL583917; CAC29684.1; -.
DR PIR; H86930; H86930.
DR Leproma; ML0176; -.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 382 AA; 37084 MW; 3D8DDBDAE32A80D CRC64;

Query Match 25.5%; Score 460; DB 16; Length 382;

Best Local Similarity 36.4%; Pred. No. 7.3e-17;
Matches 130; Conservative 52; Mismatches 121; Indels 54; Gaps 12;
QY 22 GLGLATAQAAPALSDQFADFPALPLDPSA---MVAQVAPQVWVNTKLGYNNA 77
Db 50 GAGPVTGAAASVPAANM-----PSGVEQVAVKVPVVMLETDLGEQSE-- 94
QY 78 AGTGVIDPQVNVLTNNHVIAGT-----DINAFSVGSGQTVGVVDVDRQDV 123
Db 95 EGSGLVLSADGLITNNHVIAAAKPPLGSPPPKTTVTFTF--DGTASFTVVGADP 151
QY 124 TQDVAVLQRLGAGLPSAIAIGGG--VAVGEPVVMWNGSGGQGTFRVPGRVWALGQTVQ 181
Db 152 TSDIAVVRVQISGLTPITMGSSADLRVQGVVAVGSPGLGAGT--VTSGIVSALNRVPS 209
QY 182 ASDSLTGAETLNGLIQFPAALOPDGSQGVVNGLGQVVGWNTA-----SDNFQLSQ 235
Db 210 TTGE--SGNQNTVLDAIQTDAAINPGNSGALVNMNAQLVGVNSAIALTGADSDAQSGSI 268
QY 236 GQGFPAIPGQMAIAQIRSGGSPVHIGTAFILGLGV--DNNNGARVORVVGSAAPASLGI 294
Db 269 GLGFAIPVDQAKRIADLSTGKA-----SHASLGQVQVNDKDLGAKIVEVWAGNAAGV 322
QY 295 SLGISTGDVITAVDGPAPINSATAMADALNGHHPGDVSVNMQTKSGGTRTGNVTLAE 351
Db 323 NAAVEKGVVTKVDDRLJSSADALVAARSKAPGDKVSLTVQDQSGSRTVQVTLGK 379

RESULT 9
Q9Z5G6 PRELIMINARY; PRT; 452 AA.
ID Q9Z5G6
AC Q9Z5G6
DT 01-MAY-1999 (TREMELrel. 10, Created)
DT 01-MAY-1999 (TREMELrel. 10, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE Putative serine protease.
GN MLCB373.28.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RL Harris D., Taylor K.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=93188700; PubMed=8446027;
RA Eiglmeyer K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT "Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae."
RL Mol. Microbiol. 7:197-206(1993).
DR EMBL; AL035500; CAB36690.1; -.
DR PIR; T45448; T45448.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 452 AA; 45130 MW; 6CA675EB0911F983 CRC64;

Query Match 25.5%; Score 460; DB 2; Length 452;
 Best Local Similarity 36.4%; Pred. No. 9e-17;
 Matches 130; Conservative 52; Mismatches 121; Indels 54; Gaps 12;

QY 22 GLGLATAPAAQAAPPALSSQDRFADFPALPLDPSAMVAQVAPVNNINTKLGYNNAVAGT 77
 DB 120 GAGFVTGPAASVFAANN-----PSGSEVQAVKVPVSMLETDLGRQSE-- 164

QY 78 AGTGIVDPNGVLTNNHVIAG-----TDINAFSVGSGQTVGVVGYDR 123
 DB 165 EGSVILSADGLITNNHVVAAKPGGPGGLSPKTVTF--DRTASFTVWGADP 221

QY 124 TDQVAVLQLRGAGLPSAIAIGG--VAVGEPVVMNGSGGGGTTPRAVGRVVALGQTVQ 181
 DB 222 TSDIAVVRVQSIISGLTPITMGSSADLRVQPVVAVGSPGLAGT--VTSGIVSALNRPVS 279

QY 182 ASDSLTGABETUNGLIQFADAAIQPGSGGPPVNGLGQVVGMTAA-----SDNPLSQG 235
 DB 280 TTGE-SGNQTVLDAQTDAINFGSGGALVNMGGOLGVNSATATLGASDAQSGSI 338

QY 236 GQGFAPIGQAMAIAQIRSGSGSPVTHIGPTAFGLGLGVVDNNGN-GARVORVVGSPAA 294
 DB 339 GLGFAIPVDQAKRIADELISTG--KATH-----ASLGQVQVATDKGTFGAKVMDVWAGGAAA 392

QY 295 SLGISTGDVITAVDGAIPINSATAMALNGHHPGDVIVSNWOTKSGGTRTGNVTLAE 351
 DB 393 NAAVPGKGVLTVDRLISSADALVAARSKAFGKXVSLTYQDQSGSRTVQVTLGK 449

RESULT 10

Q9FBK9 ID Q9FBK9 PRELIMINARY; PRT; 542 AA.

AC Q9FBK9

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Putative protease.

GN SC05149 OR SCP8.12.

OS Streptomyces coelicolor.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Saunders D., Harris D.;

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;

RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,

RA Kinashi H., Hopwood D.A.;

RT "A set of ordered cosmids and a detailed genetic and physical map for

RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";

RL Mol. Microbiol. 21:77-96(1996).

RN [4]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2) / M145;

RC MEDLINE=21996410; PubMed=12000953;

RA Bentley S.D., Chater K.P., Cerdano-Tarraga A.-M., Challis G.L.,

RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,

RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,

RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,

RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,

RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,

RA Seeger K., Saunders D., Sharp S., Squares R., Taylor K.,

RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,

RA Hopwood D.A.;

RT "Complete genome sequence of the model actinomycete Streptomyces

RL coelicolor A3(2).";

RL Nature 417:141-147(2002).

DR EMBL:AL933122; CAC01350.1; --

DR GO: GO:0008233; F:peptidase activity; IEA.

DR GO: GO:0004295; F:trypsin activity; IEA.

DR GO: GO:0007242; P:intracellular signaling cascade; IEA.

DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.

DR InterPro: IPR009003; Cys_Ser_trypsin.

DR InterPro: IPR001478; PDZ.

DR InterPro: IPR001254; Peptidase_S1.

DR InterPro: IPR001940; Peptidase_S1C.

DR Pfam: PF00595; PDZ; 1.

DR Pfam: PF00089; trypsin; 1.

DR PRINTS: PRO0834; PROTEASES2C.

DR SMART: SM00228; PDZ; 1.

DR PROSITE: PS0106; PDZ; 1.

DR Hydroxylase; Protease; Serine protease; Complete proteome.

KW SEQUENCE 542 AA; 53761 MW; 620F7D889DF7212D CRC64;

SQ

Query Match 22.6%; Score 407; DB 16; Length 542;
 Best Local Similarity 32.6%; Pred. No. 7e-14;
 Matches 116; Conservative 56; Mismatches 138; Indels 46; Gaps 12;

QY 22 GLGLATAPAAQAAPPALSSQDRFADFPALPLDPSAMVAQVAPVNNINTKLGYNNAVAGT 81
 DB 206 GVGTVELP-QAGPEAAERD-----PDSVAGTAARALPSVTLH--VSGSEAGTGTG 254

QY 82 IVIDPENGVLITNNHVI--AGATDINAFSVGSGQTVGVVGYDRVQVAVLQLRGAGLPL 139
 DB 255 FVLDCRGHILITNNHVVPEAGSGEITVTFNSGDTAAEAEEVWGRDGGYLVAVKVGTVGLT 314

QY 140 SAAIGG--GVAVGEPVWAMGSGGGTTPRAVGRVVALGQTVQVQ--SDSLTGABETUNG 196
 DB 315 PMPGLNSDNRVGVDFVVALGAPFLAGT--VTSGIIIAKRPITAGGEGDGSISYVDA 372

QY 197 IQFDDAAIQPGSGGPPVNGLGQVVGMTA-----ASDNFOLSGGGGPAIPICQAM 247
 DB 373 LQTDAPINFGSGGLDARGAIGNSAIRSADSGSTESDDGQAGSIGLGFAPINQKG 432

QY 248 ALAGAIRSGGSPVTHIGPTAFGLGV-VDNN--GNGARVQ-----RVVGSAPAA 296
 DB 433 RVABEL-----INTCKAHPVIGITLDNNTYGDGARVSAGKGGDGPVTTGGP 484

QY 297 GISTGDVITAVDGAIPINSATAMALNGHHPGDVIVSNWOTKSGGTRTGNVTLAE 352
 DB 485 GIKPGDVITAVDQVRVHSGEELIVKTRAHPGDRLELTQDRGKETKVSIVLSSG 540

RESULT 11

Q93J30 ID Q93J30 PRELIMINARY; PRT; 519 AA.

AC Q93J30

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Putative protease.

GN SC03977 OR SCBAC35E3.14.

OS Streptomyces coelicolor.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycineae; Streptomycetaceae; Streptomyces.

OX NCBI_TaxID=1902;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Collins M., Harris D.;

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=A3(2);

RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;

RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.

```

[3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaita D., Eichner A., Cullum J.,
RA Kiraishi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Bogle A., Hidalgo J., Hornsby T., Howarth S.,
RA Cronin A., Fraser A., Gowle A., Harkiss J., Oliver K., O'Neill S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Rutter S.,
RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S., Taylor K.,
RA Seeger K., Saunders D., Sharp S., Squares S., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
DR EMBL; AL939118; CAC4701.1; -.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00202; Tryp_Spc; 1.
DR SMART; SM00202; Tryp_Spc; 1.
KW Hydrolyase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 519 AA; 50327 MW; 5CB8D5F0CC19E428 CRC64;

Query Match 22.0%; Score 396; DB 16; Length 519;
Best Local Similarity 32.2%; Pred. No. 2.5e-13;
Matches 119; Conservative 54; Mismatches 148; Indels 48; Gaps 13;

QY 14 LLSVLAAGVLG-----LATAPQAAPPALSDQRFADFPALPLDPSAMVAQA 60
DB 159 LIAALVAGLGGGLGYTLARNDBSGSTTVSASDTGGSKVRDACTVA-----GVAAKAL 212

QY 61 PQVNTNKLGYNNVAGNGTGVDPNGVLTNNHVIAGATDINAFSV--GSGOTYGV 118
DB 213 PSTVTIQAB-GSNGEGGTGTGFVDFKEGHIIVNNHVAEAVDGGKLSATFPNGKKYDAEV 271

QY 119 VGYDRTQDVAVLQRLGA--GGLPSAAIGGG--VAYGEFVAMGNSGGGGTTPRAVPGRWVA 175
DB 272 VGHAGQYDVAVIKLENAPSLDKPLALGDSDKVAVGDSITAIAPFGLSNT--VTTGIIISA 329

QY 176 LGQTVQASDLSLTGAETLNGLIQFDAAIQDSDSGPVNGLGVVGNWNTA-----ASDNFQ 231
DB 330 KNRFPVSSDGSADSKASTMSALQTDASINPNSGGLDADQGNVIGINSIQSTGNGGFG 389

QY 232 LSCQSGQ--GFAIPIGQAMAIAGQIRSGGSGPTVHIGTAPFLGIGVDDNNGGARV--QR 286
DB 390 TQQAGSIGLGFAPVNOAKFVAQQLIKSGKRVYAKIGASVSL-----EETNGAKLTEQG 444

QY 287 VVGS-----APASLGLSTGCVITAVDGAIPINSATAMADALNCHHFGDVISVNWQYKSG 340
DB 445 VGGSDPVEKGGFADAGLKPGLDITKLLDRVIDSGPTLIGETIWTHTKPGDVEVTYIE-RGG 503

QY 341 GTRIGNVTIL 349
DB 504 KQHTAEVTL 512

```

```

RESULT 12
Q8DG87 PRELIMINARY; PRT; 375 AA.
AC Q8DG87;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Serine protease.
GN TLR2436.
OS Synecococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synecococcus.
OX NCBI_TaxID=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BP-1;
RX MEDLINE=22225144; PubMed=12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriiguchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimpō S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1.";
RL DNA Res. 9:123-130(2002).
DR EMBL; AF053377; BAC09988.1; -.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 375 AA; 39724 MW; 0D2602EB534915F5 CRC64;

Query Match 21.7%; Score 391; DB 16; Length 375;
Best Local Similarity 33.6%; Pred. No. 3.1e-13;
Matches 108; Conservative 49; Mismatches 114; Indels 50; Gaps 10;

QY 35 PALSDQRFAD-FPALPLDPSAMVAQVAVQVNVNTKLGYNNVAGNGTGVDPNGVVLTN 93
DB 67 PLLSDPFFRFQFFPGLALPQE-----DLRGGSGFLIDPSGIVMIN 108

QY 94 NEVIAGATDINAFSGSGTGVGVYDRTQDVAVLQRLG-AGGLPSAAIGGG--VAYG 150
DB 109 AHVVSQADTVNV-RLKDGSRVFEVGVDEVSDLAIVLKGVTEPLTAPLGDSEVKVG 167

QY 151 EFPVAMNSGGGGTTPRAVPGRWVAG--QTVQASDLSLTGAETLNGLIQFDAAIQDSD 208
DB 168 DWAIAGVNPGLDNT-----VTLGIISTLHRSAAQVGIPIKRLDFIQTDAAINPGNS 219

QY 209 GGFVNVNGLGVVGMNTAASDNFQLSGGGGFAIPICQAMAIAGQIRSGGSGPTVHIG--P 266
DB 220 GGPLNNEAGEVIGINTA-----IRADWNGIGFAIPINKAKALQARLIRGEKIYHAYIGQM 275

QY 267 TAPFLGVLGVVDNNG-----NGARVQRVVGSGAPASLGIISTGCVITAVDGAIPINSA 315
DB 276 TTFTPAMAXENNANPNPSPVILPEVNVGLVQLVPLNTPAKAGLRWGDVITAVDGEPI TSA 335

QY 316 TAMADALNCHHFGDVISVNWQ 336
DB 336 DQLQTTIVDSAAVGVQVNLNLTVQ 356

RESULT 13
P72780

```



```
QY 275 VDNNGGARVQVRVGGAPAAISLGISTGDTVAVDGPAINSATAMADALNGHHPGDVISVN 334
Db 394 TEOGASGS--DAITPNPRAKAGLKPGDVTIKLDWVIDSGTFLIGEINWTRPGATVKLT 451
QY 335 WQTKSGGTGNGVTLAE 351
Db 452 Y-TRDGKARTDVTLGE 467

RESULT 15
Q821L8 PRELIMINARY; PRT; 619 AA.
AC Q821L8;
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Putative serine proteinase.
GN SAV3115.
OS Streptomyces avermitilis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=33903;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=21477403; PubMed=11572948;
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA Shinoe M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.,
RT "Genome sequence of an industrial microorganism Streptomyces
RT avermitilis: deducing the ability of producing secondary
RT metabolites."
RL Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;
RX MEDLINE=22608305; PubMed=12625252;
RA Ikeda H., Ishikawa J., Hanamoto A., Shinoe M., Kikuchi H., Shiba T.,
RA Sakaki Y., Hattori M., Omura S.;
RT "Complete genome sequence and comparative analysis of the industrial
RT microorganism Streptomyces avermitilis."
RL Nat. Biotechnol. 21:526-531(2003).
DR EMBL: AP005033; BAC70826.1;
DR GO: GO:0004295; F:trypsin activity; IEA.
DR GO: GO:0007242; P:intracellular signaling cascade; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR009003; Cys_Ser_trypsin.
DR InterPro: IPR01478; PDZ.
DR InterPro: IPR01254; Peptidase S1.
DR InterPro: IPR01940; Peptidase_SIC.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ; 1.
DR Complete proteome.
SQ SEQUENCE 619 AA; 64251 MW; D2D3C72A67D3B3A6 CRC64;
```

```
Query Match 21.2%; Score 382; DB 16; Length 619;
Best Local Similarity 33.7%; Pred. No. 1.7e-12;
Matches 114; Conservative 53; Mismatches 127; Indels 44; Gaps 12;

QY 38 SODRPAFPALPLDPSAMVAQVAVQVYNINTKLYNNNAVAGAGTIVDPNGVLTNNHVI 97
Db 298 SEERAADSV-----GIAARALPSVTLHVK--GSAEGTGTGFLDGRGHILTNNHV 349
QY 98 --AGATDINAFVSGQTYGVVVGYDRTQDVAVLQRLGAGLPSAAIGG--GVAVGEFV 153
Db 350 EPAGSSGEISVTFSGGETAKATVVRGDSGYDLAVVKGVGGLKPMPLGNSDNVQGDV 409
QY 154 VAMGNSGGGGTTPRAVGRVVALGQTVQA--SDSLTGAETLNGLIQFDAAIQGDSGGPV 212
Db 410 VAIGAPFDLANT--VTSGLISAKERPITAGGKGGSDSVYVDALQTDAPINFGNSGGL 467
```

```
QY 213 VNLGQVVMNTAA-----SNFQLSQGG--GFAIPGQAMAIAQIRSGGSPVTHIG 265
Db 468 LDSKARVVGINSAIRSDSSDQSGAGSICLGFAPVNOAKRVAEELINTGRATHPVIG 527
QY 266 PTAFLGLGVVDNNGNGARV-----QRVVGSAPAASLGISTGDTVAVDGPAINGATA 317
Db 528 VTLDM-----DYTGEGARVGTKSNDGSPVTRGFGDRAIGQAGDVITEVDGQRIHSGE 582
QY 318 MADALNGHHPGDVISVNMQTKSGGTRTGN---VTLAEG 352
Db 583 LIVKIRAHPRGDRALATVE-----RDGKEKPVTLVLG 614
```

Search completed: June 30, 2004, 16:52:44
Job time : 30.8569 secs

1. The first part of the document is a letter from the President of the United States to the Congress, dated January 3, 1862. The letter is signed by Abraham Lincoln and is addressed to the Senate and House of Representatives. The letter discusses the state of the Union and the progress of the war against the Confederacy. It also mentions the Emancipation Proclamation and the importance of the Union's cause.

2. The second part of the document is a report from the Secretary of the War Department, dated January 3, 1862. The report is signed by Edwin M. Stanton and is addressed to the President. The report discusses the military situation and the progress of the war. It also mentions the Emancipation Proclamation and the importance of the Union's cause.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:39:00 ; Search time 40.2057 Seconds
(without alignments)
2747.774 Million cell updates/sec

Title: US-09-597-796c-8
Perfect score: 1949
Sequence: 1 MVDGALPEINSARMYAGP.....SGVLVPPRPVYVMPHSPAAG 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseq_29Jan04:*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1949	100.0	391	2	AAW32381	Aaw32381 Mycobacte
2	1949	100.0	391	2	AAW32449	Aaw32449 Mycobacte
3	1949	100.0	391	2	AAW64335	Aaw64335 Mycobacte
4	1949	100.0	391	2	AAW81702	Aaw81702 M. tuberc
5	1949	100.0	391	2	AAW04778	Aay04778 Mycobacte
6	1949	100.0	391	2	AAW38989	Aay38989 M. tuberc
7	1949	100.0	391	2	AAW39132	Aay39132 M. tuberc
8	1949	100.0	391	4	AAU01888	Aau01888 M. tuberc
9	1949	100.0	391	5	AAE29707	Aae29707 Mycobacte
10	1949	100.0	391	5	AAE17571	Aae17571 Mycobacte
11	1949	100.0	596	2	AAW32070	Aay32070 Mycobacte
12	1949	100.0	596	5	AAE29710	Aae29710 Mycobacte
13	1949	100.0	596	5	AAE17574	Aae17574 Mycobacte
14	1949	100.0	599	5	AAU74599	Aau74599 Antigenic
15	1949	100.0	600	5	AAW32068	Aay32068 Mycobacte
16	1949	100.0	600	5	AAU74597	Aau74597 Antigenic
17	1949	100.0	723	7	ADA26354	Ada26354 Mycobacte
18	1949	100.0	729	4	AAO22142	Aao22142 Ra12-H9-3
19	1949	100.0	729	5	AAE29709	Aae29709 Mycobacte
20	1949	100.0	729	5	AAE17573	Aae17573 Mycobacte
21	1949	100.0	729	7	ADA26374	Ada26374 Mycobacte
22	1949	100.0	744	4	AAU01902	Aau01902 M. tuberc
23	1949	100.0	813	7	ADA26367	Ada26367 Mycobacte
24	1949	100.0	815	4	AAU01904	Aau01904 M. tuberc
25	1949	100.0	825	7	ADA26366	Ada26366 Mycobacte

26	1949	100.0	875	7	ADA26365	Ada26365 Mycobacte
27	1949	100.0	930	5	AAE29731	Aae29731 Mycobacte
28	1949	100.0	930	7	ADA26364	Ada26364 Mycobacte
29	1949	100.0	1010	7	ADA26356	Ada26356 Mycobacte
30	1949	100.0	1016	7	ADA26370	Ada26370 M. bovis
31	1949	100.0	1022	7	ADA26369	Ada26369 Mycobacte
32	1949	100.0	1154	7	ADA26368	Ada26368 Mycobacte
33	1945	99.8	788	4	AAU01903	Aau01903 M. tuberc
34	1944	99.7	394	2	AAU04779	Aay04779 Mycobacte
35	1944	99.7	729	5	AAE29708	Aae29708 Mycobacte
36	1944	99.7	729	5	AAE17572	Aae17572 Mycobacte
37	1939	99.5	729	7	ADA26373	Ada26373 Mycobacte
38	1931	99.1	729	5	AAU32059	Aay32059 Mycobacte
39	1902.5	97.6	726	5	AAU74588	Aau74588 Antigenic
40	1652.5	84.8	396	2	AAW64337	Aaw64337 Mycobacte
41	1652.5	84.8	396	2	AAW81704	Aaw81704 M. tuberc
42	1652.5	84.8	396	2	AAW38991	Aay38991 M. tuberc
43	1652.5	84.8	396	2	AAW39134	Aay39134 M. tuberc
44	1583	81.2	393	6	ABU36968	Abu36968 Protein e
45	1583	81.2	393	6	ABU34531	Abu34531 Protein e

ALIGNMENTS

RESULT 1
AAW32381
ID AAW32381 standard; protein; 391 AA.
XX
AC AAW32381;
XX
DT 13-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen T5H-9FL.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
XX skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
PN WO9709429-A2.
XX
PD 13-VAR-1997.
XX
PF 30-AUG-1996; 96WO-US014675.
XX
PR 01-SEP-1995; 95US-00523435.
PR 22-SEP-1995; 95US-00532136.
PR 22-VAR-1996; 96US-00620280.
PR 05-JUN-1996; 96US-00658800.
PR 12-JUL-1996; 96US-00680573.
XX
(CORI-) CORIXA CORP.
PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedwick TH, Twardzik DE;
XX
DR WPI; 1997-192904/17.
DR N-PSDB; AAT91455.
XX
PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
XX useful for diagnosis of M. tuberculosis infection.
PS Example 3; Page 150-152; 190pp; English.
XX
CC A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
CC variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a M.tuberculosis antigen,
CC T5H-9FL The immunogenic polypeptide can be used to diagnose
CC M.tuberculosis infection by forming complexes with specific antibodies in
CC the sample. Fragments of DNA encoding the immunogenic polypeptide can be
CC used as diagnostic primers or probes and agents that bind to the antigen,

CC especially monoclonal antibodies or equivalent polyclonal antibodies, are
CC also used for diagnosis

XX SQ Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 2; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.2e-142;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAAQWWDVSASDLFSAASAFQSVVWGLTVGWSIG 60
DB 1 MVDGALPPEINSARMYAGPGSASLVAAAQWWDVSASDLFSAASAFQSVVWGLTVGWSIG 60
QY 61 SSAGLMVAAAAPYVAMSVTGAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAAAAPYVAMSVTGAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGONTPTAIIVNEAEYGEWMAQDAAMFGVAAAATATATATLPPFEAPEMTSAGG 180
DB 121 LIATNLLGONTPTAIIVNEAEYGEWMAQDAAMFGVAAAATATATATLPPFEAPEMTSAGG 180
QY 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGGLWKTVPSPHSPI 240
DB 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGGLWKTVPSPHSPI 240
QY 241 MYSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAAQAVTPAARALPLTSLTSAERPGQMLGGLPV 300
DB 241 MYSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAAQAVTPAARALPLTSLTSAERPGQMLGGLPV 300
QY 301 LGGGVAANLGRAASVGSLSVPOQAWAANQAVTPAARALPLTSLTSAERPGQMLGGLPV 360
DB 301 LGGGVAANLGRAASVGSLSVPOQAWAANQAVTPAARALPLTSLTSAERPGQMLGGLPV 360
QY 361 GQMGARAGGGLSGVLVPPRPVPMHPSAAG 391
DB 361 GQMGARAGGGLSGVLVPPRPVPMHPSAAG 391

RESULT 2
AAW32449

ID AAW32449 standard; protein; 391 AA.

XX AC AAW32449;

XX DT 09-JAN-1998: (first entry)

DE Mycobacterium tuberculosis antigen TbH-9PL.

XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
KW skin testing; M.tuberculosis.

XX OS Mycobacterium tuberculosis.

XX FN WO9709428-A2.

XX PD 13-MAR-1997.

XX PF 30-AUG-1996; 96WO-US014674.

XX PR 01-SEP-1995; 95US-00523436.

XX PR 22-SEP-1995; 95US-00533634.

XX PR 22-MAR-1996; 96US-00620874.

XX PR 05-JUN-1996; 96US-00659683.

XX PR 12-JUL-1996; 96US-00680574.

XX PA (CORI-) CORIXA CORP.

XX PI Reed SG, Skeiky YA, Dillon DC, Campos-Neto A, Houghton R;

XX PI Vedwick TH, Twardzik DR;

XX DR WPI; 1997-192903/17.

XX DR N-PSDB; AAT91521.

PT New immunogenic polypeptide(s) from Mycobacterium tuberculosis - are
PT useful in vaccines for prevention or treatment of tuberculosis, also for
PT diagnosis.

XX Example 3; Page 138-139; 168pp; English.

XX A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
CC variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a M.tuberculosis antigen,
CC TbH-9PL. The immunogenic protein, and fusion proteins containing one or
CC more of the proteins or one of the proteins plus ESAT-6, are useful in
CC vaccines, preferably when formulated with a non-specific adjuvant, to
CC induce an immune response against M.tuberculosis (for treatment or
CC prevention)

XX SQ Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 2; Length 391;

Best Local Similarity 100.0%; Pred. No. 1.2e-142;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAAQWWDVSASDLFSAASAFQSVVWGLTVGWSIG 60

DB 1 MVDGALPPEINSARMYAGPGSASLVAAAQWWDVSASDLFSAASAFQSVVWGLTVGWSIG 60

QY 61 SSAGLMVAAAAPYVAMSVTGAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120

DB 61 SSAGLMVAAAAPYVAMSVTGAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120

QY 121 LIATNLLGONTPTAIIVNEAEYGEWMAQDAAMFGVAAAATATATATLPPFEAPEMTSAGG 180

DB 121 LIATNLLGONTPTAIIVNEAEYGEWMAQDAAMFGVAAAATATATATLPPFEAPEMTSAGG 180

QY 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGGLWKTVPSPHSPI 240

DB 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGGLWKTVPSPHSPI 240

QY 241 MYSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAAQAVTPAARALPLTSLTSAERPGQMLGGLPV 300

DB 241 MYSMAHNMHMTNSGVSMNTLSSMLKGFAPAAAAQAVTPAARALPLTSLTSAERPGQMLGGLPV 300

QY 301 LGGGVAANLGRAASVGSLSVPOQAWAANQAVTPAARALPLTSLTSAERPGQMLGGLPV 360

DB 301 LGGGVAANLGRAASVGSLSVPOQAWAANQAVTPAARALPLTSLTSAERPGQMLGGLPV 360

QY 361 GQMGARAGGGLSGVLVPPRPVPMHPSAAG 391

DB 361 GQMGARAGGGLSGVLVPPRPVPMHPSAAG 391

RESULT 3

AAW64335

ID AAW64335 standard; protein; 391 AA.

XX AC AAW64335;

XX DT 17-OCT-2003 (revised)

XX DT 09-NOV-1998 (first entry)

DE Mycobacterium tuberculosis antigen TbH-9PL.

XX Tuberculosis; infection; diagnosis; antigen; TbH-9PL.

XX OS Mycobacterium tuberculosis; strain H37Rv.

XX EN WO9816645-A2.

XX PD 23-APR-1998.

XX PF 07-OCT-1997; 97WO-US018214.

XX PR 11-OCT-1996; 96US-00729622.

PR 13-MAR-1997; 97US-00818111.
 XX (CORI-) CORIXA CORP.
 XX
 XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 XX Vedvick TS, Twardzik DR, Lodes MJ;
 XX WPI; 1998-251292/22.
 DR N-PSDB; AAV44395.
 XX
 XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and
 PT diagnosis of tuberculosis.
 XX
 XX Example 3; Page 133-135; 250pp; English.
 XX
 CC This polypeptide comprises Mycobacterium tuberculosis antigen TBH-9FL. It
 CC is encoded by genomic DNA (see AAV44395) isolated from a M. tuberculosis
 CC strain H37RV genomic library using a probe from clone TBH-9 (see
 CC AAV44371). The invention relates to compositions and methods for
 CC diagnosing tuberculosis. It provides polypeptides (see AAV64291-W64379)
 CC comprising an antigenic portion of a soluble M. tuberculosis antigen, or
 CC an immunogenic portion of an M. tuberculosis antigen, as well as DNA
 CC sequences encoding such polypeptides, recombinant expression vectors and
 CC transformed or transfected host cells. Also claimed are methods and
 CC diagnostic kits for detecting M. tuberculosis infection in a patient
 CC using these polypeptides, antibodies or oligonucleotide probes and
 CC primers, for the diagnosis of tuberculosis. (Updated on 17-OCT-2003 to
 CC standardise OS field)
 XX
 SQ Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 2; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142; Indels 0; Gaps 0;
 Matches 391; Conservative 0; Mismatches 0;

QY 1 MVDGALPPEINARMYAGPGSASLVAAQWMDVSADLFSAAAFQSVVWGLTVGSGWIG 60
 DB 1 MVDGALPPEINARMYAGPGSASLVAAQWMDVSADLFSAAAFQSVVWGLTVGSGWIG 60
 QY 61 SSAGLMVAAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
 DB 61 SSAGLMVAAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180
 DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATLPPPEAPMTSAGG 180
 QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQTTGTPSSKLGGLWKTVPSPHSPI 240
 DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQTTGTPSSKLGGLWKTVPSPHSPI 240
 QY 241 MVSMAHNMSTNGSVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSG 300
 DB 241 MVSMAHNMSTNGSVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSG 300
 QY 301 LGGGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
 DB 301 LGGGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
 QY 361 GQMGARAGGSLGVLRVPPRPYVMPHSPAAG 391
 DB 361 GQMGARAGGSLGVLRVPPRPYVMPHSPAAG 391

RESULT 4

AAW81702 ID AAW81702 standard; protein; 391 AA.

XX AC AAW81702;

XX DT 27-JAN-1999 (first entry)

XX

DE M. tuberculosis immunogenic polypeptide TBH-9FL.
 XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 KW vaccine; pharmaceutical; infection; diagnosis.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN WO9816646-A2.
 XX
 PD 23-APR-1998.
 XX
 PF 07-OCT-1997; 97WO-US018293.
 XX
 PR 11-OCT-1996; 96US-00730510.
 PR 13-MAR-1997; 97US-00818112.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX WPI; 1998-261042/23.
 DR N-PSDB; AAV64503.
 XX
 CC Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
 CC develop products for the detection of M. tuberculosis infection and for
 CC diagnosis, treatment and prevention of tuberculosis.
 PT
 XX Example 3B; Page 128-129; 230pp; English.
 XX
 CC This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for
 CC inducing protective immunity against tuberculosis (TB). This sequence can
 CC be formulated into vaccines and/or pharmaceutical compositions for
 CC immunising against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis
 XX
 SQ Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 2; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142; Indels 0; Gaps 0;
 Matches 391; Conservative 0; Mismatches 0;

QY 1 MVDGALPPEINARMYAGPGSASLVAAQWMDVSADLFSAAAFQSVVWGLTVGSGWIG 60
 DB 1 MVDGALPPEINARMYAGPGSASLVAAQWMDVSADLFSAAAFQSVVWGLTVGSGWIG 60
 QY 61 SSAGLMVAAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
 DB 61 SSAGLMVAAASPYVAVMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATLPPPEAPMTSAGG 180
 DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATLPPPEAPMTSAGG 180
 QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQTTGTPSSKLGGLWKTVPSPHSPI 240
 DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQTTGTPSSKLGGLWKTVPSPHSPI 240
 QY 241 MVSMAHNMSTNGSVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSG 300
 DB 241 MVSMAHNMSTNGSVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSG 300
 QY 301 LGGGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
 DB 301 LGGGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
 QY 361 GQMGARAGGSLGVLRVPPRPYVMPHSPAAG 391
 DB 361 GQMGARAGGSLGVLRVPPRPYVMPHSPAAG 391

RESULT 5

AAV04778
 ID AAY04778 standard; protein; 391 AA.
 XX AC AAY04778;
 XX DT 06-JUL-1999 (first entry)
 XX DE Mycobacterium species protein sequence 5R.
 XX KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 XX KW hybridisation; detection; vaccine; immunisation; infection.
 XX OS Mycobacterium sp.
 XX PN WO9909186-A2.
 XX PD 25-FEB-1999.
 XX PF 14-AUG-1998; 98WO-FR001813.
 XX PR 14-AUG-1997; 97FR-00010404.
 XX PR 11-SEP-1997; 97FR-00011325.
 XX PA (INSP) INST PASTEUR.
 XX PI Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;
 XX PI Goguet De La Salmoniere Y;
 XX DR WPI: 1999-181045/15.
 XX DR N-PSDB; AAX34030.
 XX PT Mycobacterial DNA vectors containing reporter constructs - for
 XX PT identifying coding or promoter sequences involved in infection-associated
 XX PT protein expression.
 XX PS Claim 32; Fig 5R; 309pp; French.
 XX CC Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
 XX CC from various Mycobacterium species microorganisms. The encoding
 XX CC nucleotide sequences can be used as primers and probes for methods for
 XX CC detecting and identifying mycobacteria, especially belonging to the M.
 XX CC tuberculosis complex. The encoded proteins can be used in vaccines for
 XX CC immunisation against a bacterial or viral infection
 XX CC
 XX CC Sequence 391 AA;
 Query Match 100.0%; Score 1949; DB 2; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MVDFGALPPEINSAARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 DB 1 MVDFGALPPEINSAARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 QY 61 SSAGLMVAAASPYVAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 DB 61 SSAGLMVAAASPYVAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPMTSAGG 180
 DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPMTSAGG 180
 QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
 DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
 QY 241 MVSMAHHMNTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
 DB 241 MVSMAHHMNTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
 QY 301 LGGVAAANLGAASVGSLSVPOAAANOAVTPAARALPLTSLTSAERPGQWGLGLPV 360
 DB 301 LGGVAAANLGAASVGSLSVPOAAANOAVTPAARALPLTSLTSAERPGQWGLGLPV 360

QY 361 GQMGARAGGSLGVLVRPVRPRPYVMPHSPAAG 391
 DB 361 GQMGARAGGSLGVLVRPVRPRPYVMPHSPAAG 391
 RESULT 6
 AAY38989
 ID AAY38989 standard; protein; 391 AA.
 XX AC AAY38989;
 XX DT 05-NOV-1999 (first entry)
 XX DE M. tuberculosis recombinant antigen protein Tbh-9FL.
 XX KW Antigen; diagnosis; detection; infection; antibody; immunisation;
 XX KW vaccine; immunity.
 XX OS Mycobacterium tuberculosis.
 XX PN WO9942118-A2.
 XX PD 26-AUG-1999.
 XX PF 17-FEB-1999; 99WO-US003265.
 XX PR 18-FEB-1998; 98US-00024753.
 XX PR 05-MAY-1998; 98US-00072596.
 XX PA (CORI-) CORIXA CORP.
 XX PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 XX PI Vedwick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
 XX DR WPI: 1999-527416/44.
 XX DR N-PSDB; AAZ19093.
 XX PT New polypeptide comprising antigenic portions of M. tuberculosis.
 XX PS Example 3; Page 168-169; 323pp; English.
 XX CC This invention describes novel recombinant antigens and their encoding
 XX CC nucleic acids derived from Mycobacterium tuberculosis. The novel
 XX CC polypeptides are useful for detecting M. tuberculosis infection in a
 XX CC biological sample by detecting antibodies which bind with the
 XX CC polypeptides, and are useful as vaccines for immunizing against M.
 XX CC tuberculosis infection. The new detection methods are needed as current
 XX CC vaccination strategies do not provide 100% immunity
 XX CC
 XX CC Sequence 391 AA;
 Query Match 100.0%; Score 1949; DB 2; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MVDFGALPPEINSAARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 DB 1 MVDFGALPPEINSAARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
 QY 61 SSAGLMVAAASPYVAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 DB 61 SSAGLMVAAASPYVAMSVTAGQBELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATLTPFEAPMTSAGG 180
 DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATLTPFEAPMTSAGG 180
 QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
 DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
 QY 241 MVSMAHHMNTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300

```

Db      241  MVSNNHMTSGVSMNTLTSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSSG 300
QY      301  LGGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
Db      301  LGGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
QY      361  GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db      361  GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391

RESULT 7
AAU01888
ID  AAU01888 standard; protein; 391 AA.
AC  AAU01888;
XX
DT  26-AUG-1999.
DE  17-FEB-1999; 99WO-US003268.
PR  18-FEB-1998; 98US-00025197.
PR  05-MAY-1998; 98US-00072967.
XX
PA  (CORI-) CORIXA CORP.
PI  Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI  Vedvick TS, Twardzik DR, Lodes MJ, Hendrickson RC;
XX
DR  WPI; 1999-527409/44.
DR  N-PSDB; AA219305.
XX
PT  New antigens from Mycobacterium tuberculosis useful in diagnostic skin
PT  tests and protective or therapeutic vaccines or compositions.
XX
PS  Example 3; Page 123-124; 299pp; English.
XX
CC  The present invention describes polypeptides comprising an immunogenic
CC  part of a Mycobacterium tuberculosis antigen (Ag). Also described are
CC  vaccines and fusion protein containing M. tuberculosis Ag's. M.
CC  tuberculosis Ag's, DNAs encoding them, derived fusion proteins and other
CC  polypeptides fragments, can be used in pharmaceutical compositions or
CC  vaccines to generate a protective or therapeutic immune response to M.
CC  tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
CC  Ag can induce proliferation of, or cytokine secretion by, T, B or natural
CC  killer cells and/or macrophages in tuberculosis-immune subjects. AA219249
CC  to AA219460 and AA219383 to AA219225 are used in the exemplification of
CC  the present invention
XX
SQ  Sequence 391 AA;

Query Match      100.0%; Score 1949; DB 2; Length 391;
Best Local Similarity 100.0%; Pred No. 1.2e-142; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 0;

QY      1  MVDFGALPPPEINSGARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
Db      1  MVDFGALPPPEINSGARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
QY      61  SSAGLMVAASPYVAMNSVTAGQAELTAAQVRVAAAAYETAYGLTVPPVTAENRAELMI 120

```

```

Db      61  SSAGLMVAASPYVAMNSVTAGQAELTAAQVRVAAAAYETAYGLTVPPVTAENRAELMI 120
QY      121  LIATNLLGQNTPATAVNEABYGENWAODAAAMFGVAAATATATATLLPFEAPBMTSAGG 180
Db      121  LIATNLLGQNTPATAVNEABYGENWAODAAAMFGVAAATATATATLLPFEAPBMTSAGG 180
QY      181  LLEQAAAVEEASDTAAANCLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
Db      181  LLEQAAAVEEASDTAAANCLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
QY      241  MVSNNHMTSGVSMNTLTSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSSG 300
Db      241  MVSNNHMTSGVSMNTLTSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSSG 300
QY      301  LGGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
Db      301  LGGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
QY      361  GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db      361  GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391

RESULT 8
AAU01888
ID  AAU01888 standard; protein; 391 AA.
AC  AAU01888;
XX
DT  29-AUG-2001 (first entry)
DE  M. tuberculosis antigen Tbh9 (Mtb39A).
XX
KW  Tbh9; Mtb39A; antigen; vaccine; tuberculosis; AIDS;
KW  acquired immunodeficiency disease.
XX
OS  Mycobacterium tuberculosis.
XX
PN  WO200124820-A1.
XX
PD  12-APR-2001.
XX
PF  10-OCT-2000; 2000WO-US028095.
PR  07-OCT-1999; 99US-0158338P.
PR  07-OCT-1999; 99US-0158425P.
XX
PA  (CORI-) CORIXA CORP.
XX
PI  Skeiky Y, Reed S, Houghton RL, Mcneill PD, Dillon DC, Lodes ML;
XX
DR  WPI; 2001-290576/30.
DR  N-PSDB; AA503779.
XX
PT  Vaccinating against Mycobacteria infections in mammals using fusion
PT  proteins comprising combinations of heterologous antigens.
XX
PS  Example 2; Page 151-152; 168pp; English.
XX
CC  The sequence represents Mycobacterium tuberculosis Tbh9 (also known as
CC  Mtb39A), an M. tuberculosis antigen. Compositions comprising at least 2
CC  heterologous antigens, as a fusion protein, and vectors expressing the
CC  fusion proteins are used as vaccines to prophylactically immunise mammals
CC  (especially humans) against infection by Mycobacteria. The compositions
CC  contain at least 2 heterologous antigens that increase the serological
CC  sensitivity of individuals infected with tuberculosis, a disease
CC  frequently affecting patients with acquired immunodeficiency disease,
CC  AIDS
XX
SQ  Sequence 391 AA;

Query Match      100.0%; Score 1949; DB 4; Length 391;

```

Best Local Similarity 100.0%; Pred. No. 1.2e-142; Mismatches 0; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 1 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60

QY 61 SSAGLMVAAAAPYVAVMSVTAGQAELETAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
DB 61 SSAGLMVAAAAPYVAVMSVTAGQAELETAQVRVAAAAYETAYGLTVPPIAENRAELMI 120

QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGG 180

QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240

QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLSGLSSG 300
DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLSGLSSG 300

QY 301 LGGVVAANLGRAASVGSLSVFPQAWAANAQVTPAARALPLTSLTSAABERGQMLGGLPV 360
DB 301 LGGVVAANLGRAASVGSLSVFPQAWAANAQVTPAARALPLTSLTSAABERGQMLGGLPV 360

QY 361 GQMGARAGGSLSGVLRVPRPYVMPHSPAAG 391
DB 361 GQMGARAGGSLSGVLRVPRPYVMPHSPAAG 391

RESULT 9
AAE29707
ID AAE29707 standard; protein; 391 AA.
XX
AC AAE29707;
XX
DT 27-JAN-2003 (first entry)
XX
DE Mycobacterium sp. TbH9FL antigenic protein.
XX
KW Vaccine; immunity; diagnostic agent; gene therapy; TbH9FL antigen.
XX
OS Mycobacterium sp.
XX
FN WO200272792-A2.
XX
PD 19-SEP-2002.
XX
PF 13-MAR-2002; 2002WO-US008223.
XX
PR 13-MAR-2001; 2001US-0275837P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Brannon M, Guderian J;
XX
DR WPI; 2002-759844/82.
DR N-P5DB; AAD47082.

New recombinant nucleic acid molecule comprising a Leishmania TSA, Leif, M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity against pathogenic microorganisms e.g. Leishmania and Mycobacterium tuberculosis.

Disclosure; Page 86-87; 155pp; English.

The invention relates to a recombinant nucleic acid molecule encoding a fusion polypeptide. The recombinant nucleic acid comprises a heterologous polynucleotide sequence encoding an antigen or an antigenic fragment from Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a polypeptide or its fragment. The Leishmania polynucleotide is selected

CC from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention are used in methods for eliciting immune response in mammals. They are useful as vaccines to elicit protective immunity against pathogenic microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion polypeptides are used for enhancing the expression of polynucleotides, as in vivo diagnostic agents and for raising antibodies in a non-human animal. The invention is used in gene therapy. The present sequence is CC Mycobacterium sp. TbH9FL antigenic protein
XX
SQ Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 5; Length 391;
Best Local Similarity 100.0%; Pred. No. 1.2e-142;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 1 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60

QY 61 SSAGLMVAAAAPYVAVMSVTAGQAELETAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
DB 61 SSAGLMVAAAAPYVAVMSVTAGQAELETAQVRVAAAAYETAYGLTVPPIAENRAELMI 120

QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGG 180

QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240
DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPISN 240

QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLSGLSSG 300
DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSSLSGLSSG 300

QY 301 LGGVVAANLGRAASVGSLSVFPQAWAANAQVTPAARALPLTSLTSAABERGQMLGGLPV 360
DB 301 LGGVVAANLGRAASVGSLSVFPQAWAANAQVTPAARALPLTSLTSAABERGQMLGGLPV 360

QY 361 GQMGARAGGSLSGVLRVPRPYVMPHSPAAG 391
DB 361 GQMGARAGGSLSGVLRVPRPYVMPHSPAAG 391

RESULT 10

AAE17571
ID AAE17571 standard; protein; 391 AA.

XX
AC AAE17571;
XX

DT 22-APR-2002 (first entry)
XX

DE Mycobacterium species MTB39 (TbH9) protein #2.

XX Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; MTB39; TbH9 protein.
XX

OS Mycobacterium sp.
XX

XX WO200198460-A2.
XX

XX 27-DEC-2001.
XX

XX 20-JUN-2001; 2001WO-US019959.
XX

XX 20-JUN-2000; 2000US-00597796.
XX

XX 01-FEB-2001; 2001US-0265737P.
XX

XX (CORI-) CORIXA CORP.
XX

XX Skeiky Y, Reed S, Alderson M;
XX

XX WPI; 2002-147798/19.
XX

DR N-PSDB; AAD28341.
 XX Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 PS Claim 83; Page 102-103; 136pp; English.
 XX
 XX The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 XX MTB39 (TbH9) protein
 XX
 SQ Sequence 391 AA;

Query Match 100.0%; Score 1949; DB 5; Length 391;
 Best Local Similarity 100.0%; Pred. No. 1.2e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
 DB 1 MVDFGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
 QY 61 SSAGLMVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 DB 61 SSAGLMVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
 DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
 QY 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQTQGTTPSSKLGGLWKTVPSPHPSISN 240
 DB 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQTQGTTPSSKLGGLWKTVPSPHPSISN 240
 QY 241 MYSMANNHMTNSGVSWTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
 DB 241 MYSMANNHMTNSGVSWTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
 QY 301 LGGVVAANLGRAASVGLSVPOQAAWAAQAVTPAARALPLTSLTSAERPGQMLGGLPV 360
 DB 301 LGGVVAANLGRAASVGLSVPOQAAWAAQAVTPAARALPLTSLTSAERPGQMLGGLPV 360
 QY 361 GQMGARAGGGLSGVLRVPPPPYVMPHSPAAG 391
 DB 361 GQMGARAGGGLSGVLRVPPPPYVMPHSPAAG 391

RESULT 11
 AAY32070
 ID AAY32070 standard; protein; 596 AA.
 XX
 XX
 AC AAY32070;
 XX
 XX 17-JAN-2000 (first entry)
 DT
 XX Mycobacterium tuberculosis antigen fusion protein Mtb59f.
 DE

XX Tuberculosis; antigen; fusion protein; Mtb59f; TbH9; Ra35; diagnosis;
 KW therapy; vaccine; immunogen.
 XX
 OS Mycobacterium tuberculosis.
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..8
 FT Protein /note= "Met/His tag"
 FT Protein 9..140
 FT Protein /note= "Ra12"
 FT Protein 143..596
 FT Protein /note= "TbH9"
 XX
 PN WO9511748-A2.
 XX
 PD 14-OCT-1999.
 XX
 PF 07-APR-1999; 99WO-US007717.
 XX
 PR 07-APR-1999; 98US-00056556.
 PR 30-DEC-1999; 98US-00223040.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 XX Skeiky YAW, Alderson M, Campos-Neto A;
 PI WPI; 1999-601610/51.
 DR N-PSDB; AAZ20205.
 XX
 XX New fusion proteins useful for diagnosis, prevention and treatment of
 PT tuberculosis.
 XX
 PS Claim 1; Fig 12A-B; 83pp; English.
 XX
 CC This sequence represents a recombinant Mycobacterium tuberculosis bi-
 CC antigen fusion protein, termed Mtb59f, composed of the antigens TbH9 and
 CC Ra35. The fusion protein is expressed in host cells using a vector
 CC carrying a polynucleotide (see AAZ20205) comprising the coding sequences
 CC for the 2 antigens. The invention provides fusion proteins (see AAY32059-
 CC 71) containing at least 2 M. tuberculosis antigens. The new fusion
 CC proteins and polynucleotides encoding them are useful as vaccines for
 CC preventing tuberculosis (claimed), for diagnosis (via in vitro assays or
 CC intradermal skin tests for detection of anti-M. tuberculosis antibodies),
 CC monitoring of disease progression, and treatment of tuberculosis. They
 CC are more effective immunogens than mixtures of the individual protein
 CC components
 XX
 SQ Sequence 596 AA;

Query Match 100.0%; Score 1949; DB 2; Length 596;
 Best Local Similarity 100.0%; Pred. No. 2.1e-142;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
 DB 9 MVDFGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 68
 QY 61 SSAGLMVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 DB 69 SSAGLMVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
 DB 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
 QY 181 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQTQGTTPSSKLGGLWKTVPSPHPSISN 240
 DB 189 LLEQAAAVEEASDTAAANQLMNNVPOALQQLAQTQGTTPSSKLGGLWKTVPSPHPSISN 248
 QY 241 MYSMANNHMTNSGVSWTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
 DB 249 MYSMANNHMTNSGVSWTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 308

QY 301 LGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALPLTSLTSAABRGQOMLGGLPV 360
 Db 309 LGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALPLTSLTSAABRGQOMLGGLPV 368
 QY 361 GOMGARAGGGLSGVLVRPVPVMPHSPAAG 391
 Db 369 GOMGARAGGGLSGVLVRPVPVMPHSPAAG 399

RESULT 12

AAE29710
 ID AAE29710 standard; protein; 596 AA.

AC AAE29710;

DT 29-AUG-2003 (revised)

DT 27-JAN-2003 (first entry)

DE Mycobacterium sp. MTB59F fusion protein.

KW Vaccine; immunity; diagnostic agent; gene therapy; TBH9; antigen; Ra35;
 KW MTB59F; fusion protein.

XX Mycobacterium sp.

OS Mycobacterium tuberculosis.

OS Chimeric.

PN WC200272792-A2.

XX 19-SEP-2002.

PF 13-MAR-2002; 2002WO-US008223.

XX 13-MAR-2001; 2001US-0275837P.

XX (CORI-) CORIXA CORP.

PI Skeiky Y, Brannon M, Guderian J;

XX WPI; 2002-759944/82.

DR N-PSDB; AAD47086.

PT New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 PT tuberculosis.

PS Disclosure; Page 98-99; 155pp; English.

XX The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is
 CC MTB59F fusion protein. This fusion protein comprises Ra35 protein from
 CC Mycobacterium tuberculosis and TBH9 protein from Mycobacterium sp.
 CC (Updated on 29-AUG-2003 to standardise OS field)

XX Sequence 596 AA;

Query Match

Best Local Similarity 100.0%; Score 1949; DB 5; Length 596;

Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60

Db 9 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
 QY 61 SSAGLVAAASPVVAMSVTAQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMI 120
 Db 69 SSAGLVAAASPVVAMSVTAQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMI 128
 QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMAFGVAATAATATATATALLPPEEAPMTSAGG 180
 Db 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMAFGVAATAATATATALLPPEEAPMTSAGG 188
 QY 181 LLEQAAAABEASDTAAANQLMNNVPAALQQLAQPTGCTTPSSKLGGLWKTVPSPHSPISN 240
 Db 189 LLEQAAAABEASDTAAANQLMNNVPAALQQLAQPTGCTTPSSKLGGLWKTVPSPHSPISN 248
 QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
 Db 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLGSSLGSSG 308
 QY 301 LGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALPLTSLTSAABRGQOMLGGLPV 360
 Db 309 LGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALPLTSLTSAABRGQOMLGGLPV 368
 QY 361 GOMGARAGGGLSGVLVRPVPVMPHSPAAG 391
 Db 369 GOMGARAGGGLSGVLVRPVPVMPHSPAAG 399

RESULT 13

AAE17574

ID AAE17574 standard; protein; 596 AA.

XX AAE17574;

DT 22-APR-2002 (first entry)

XX Mycobacterium species MTB59F fusion protein.

KW Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB59F; TBH9-Ra35 protein.

XX Mycobacterium sp.

XX WO200198460-A2.

XX 27-DEC-2001.

XX 20-JUN-2001; 2001WO-US019959.

XX 20-JUN-2000; 2000US-00597796.

XX 01-FEB-2001; 2001US-0265737P.

XX (CORI-) CORIXA CORP.

XX Skeiky Y, Reed S, Alderson M;

XX WPI; 2002-147798/19.

XX N-PSDB; AAD28344.

PT Composition comprising MTB39 antigen and MTB32A antigen from
 PT Mycobacterium species, useful for eliciting immune response in a subject.
 PS Claim 5; Page 114-115; 136pp; English.

XX The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,

immunised with BCG. They are useful in the diagnosis, treatment and prevention of Mycobacterium infection. The fusion proteins and the polynucleotides are useful as diagnostic tools in patients infected with Mycobacterium, in vitro and in vivo assays for detecting humoral antibodies or cell-mediated immunity against M. tuberculosis, for the diagnosis of an infection or monitoring of disease progression, as immunogens to generate or elicit a protective immune response in a patient and for raising anti-M. tuberculosis antibodies in a non-human animal. Sequences of the invention are also used as vaccines. MTB32A fusion proteins of the invention are useful as in vivo diagnostic agents for intradermal skin test. The present sequence is Mycobacterium species MTB59f (Tb59-Ra35) fusion protein

XX Sequence 596 AA;
Query Match 100.0%; Score 1949; DB 5; Length 596;
Best Local Similarity 100.0%; Pred. No. 2.1e-142;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 60
Db 9 MVDGALPPEINSARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 68
QY 61 SSAGLWVAASPYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTPPPVIAENRAELMI 120
Db 69 SSAGLWVAASPYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTPPPVIAENRAELMI 128
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPMTSAGG 180
Db 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPMTSAGG 188
QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTGTTTSSKLGGLWKTVPSPHSPISN 240
Db 189 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTGTTTSSKLGGLWKTVPSPHSPISN 248
QY 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
Db 249 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 308
QY 361 GQMGARAGGSLGVLRVPPRPYVMPHSPAAG 391
Db 369 GQMGARAGGSLGVLRVPPRPYVMPHSPAAG 399

RESULT 14
AAU74599
ID AAU74599 standard; protein; 599 AA.
AC AAU74599;
XX
XX
DT 29-AUG-2003 (revised)
DT 08-MAY-2002 (first entry)
XX
XX
DE Antigenic fusion protein Tb59-Ra35 (Mtb59f).
XX
XX
KW Fusion protein; tuberculosis; Mycobacterium tuberculosis;
KW tuberculostatic; immunogen; vaccine; Tb59-Ra35; Mtb59f.
XX
XX
OS Mycobacterium tuberculosis.
OS Chimeric.
XX
XX
FH Key Location/Qualifiers
FT Misc-difference 597
FT /label= OTHER
FT /note= "OTHER= Xaa. Xaa= In frame stop codon"
XX
XX
XX US2002009459-A1.
XX
XX 24-JAN-2002.

XX 07-APR-1999; 99US-00287849.
XX 13-MAR-1997; 97US-00818112.
XX 01-OCT-1997; 97US-00942578.
XX 18-FEB-1998; 98US-00025197.
XX 07-APR-1998; 98US-00056556.
XX 30-DEC-1998; 98US-00223040.
XX (REED/) REED S G.
XX PA (SKEI/) SKEIKY Y A.
XX PA (DILL/) DILLON D C.
XX PA (ALDE/) ALDERSON M.
XX PA (CAMF/) CAMPOS-NETO A.
XX
XX Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
XX WPI; 2002-171134/22.
XX N-PSDB; ASK14139.
XX
XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for
XX diagnosing, treating or preventing M. tuberculosis infection,
XX particularly as vaccine for treating or preventing tuberculosis.
XX
XX Claim 1; Fig 12; 62pp; English.
XX
XX The invention relates to a purified polypeptide which induces an immune
XX response of Mycobacterium tuberculosis. Polypeptides of the invention are
XX useful for diagnosing, treating or preventing M. tuberculosis infection,
XX particularly tuberculosis infection. In particular, the polypeptides are
XX useful as a vaccine formulation with an adjuvant to afford long-term
XX protection in animals against the development of tuberculosis. The
XX protein coding sequence may be used to encode a protein product for use
XX as an immunogen to induce and/or enhance an immune response to M.
XX tuberculosis. This sequence represents an M. tuberculosis fusion protein
XX of the invention. (Updated on 29-AUG-2003 to standardise OS field)
XX
XX Sequence 599 AA;
Query Match 100.0%; Score 1949; DB 5; Length 599;
Best Local Similarity 100.0%; Pred. No. 2.1e-142;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 60
Db 9 MVDGALPPEINSARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 68
QY 61 SSAGLWVAASPYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTPPPVIAENRAELMI 120
Db 69 SSAGLWVAASPYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTPPPVIAENRAELMI 128
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPMTSAGG 180
Db 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPMTSAGG 188
QY 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTGTTTSSKLGGLWKTVPSPHSPISN 240
Db 189 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTGTTTSSKLGGLWKTVPSPHSPISN 248
QY 241 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
Db 249 MVSMAHNMSTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 308
QY 361 GQMGARAGGSLGVLRVPPRPYVMPHSPAAG 391
Db 369 GQMGARAGGSLGVLRVPPRPYVMPHSPAAG 399

RESULT 15

AAV32068

ID AAV32068 standard; protein; 600 AA.

XX AC AAV32068;

XX DT 17-JAN-2000 (first entry)

XX DE Mycobacterium tuberculosis antigen fusion protein Mtb61f.

XX KW Tuberculosis; antigen; fusion protein; Mtb61f; TBH9; DPV; MTL; diagnosis; therapy; vaccine; immunogen.

XX OS Mycobacterium tuberculosis.

XX PN WO9951748-A2.

XX PD 14-OCT-1999.

XX PF 07-APR-1999; 99WO-US007717.

XX PR 07-APR-1998; 98US-00056556.

XX PR 30-DEC-1998; 98US-00223040.

XX PA (CORI-) CORIXA CORP.

XX PI Skeiky YAW, Alderson M, Campos-Neto A;

XX DR WPI; 1999-601610/51.

XX DR N-PSDB; AAZ20203.

XX PT New fusion proteins useful for diagnosis, prevention and treatment of tuberculosis.

XX PS Claim 1; Fig 10A-B; 83pp; English.

XX CC This sequence represents a recombinant Mycobacterium tuberculosis tri-
 antigen fusion protein, termed Mtb61f, composed of the antigens TbH9, DPV
 and MTL. The fusion protein is expressed in host cells using a vector
 carrying a polynucleotide (see AAZ20203) comprising the coding sequences
 for the 3 antigens. The invention provides fusion proteins (see AAY32059-
 71) containing at least 2 M. tuberculosis antigens. The new fusion
 proteins and polynucleotides encoding them are useful as vaccines for
 preventing tuberculosis (claimed), for diagnosis (via in vitro assays or
 intradermal skin tests for detection of anti-M. tuberculosis antibodies),
 monitoring of disease progression, and treatment of tuberculosis. They
 are more effective immunogens than mixtures of the individual protein
 components

SQ Sequence 600 AA;

Query Match 100.0%; Score 1949; DB 2; Length 600;

Best Local Similarity 100.0%; Pred. No. 2.1e-142; Mismatches 0; Indels 0; Gaps 0;

Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSARMYAGPGASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60

DB 9 MVDFGALPPEINSARMYAGPGASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 68

QY 61 SSAGIMVAASPYVWWSVTAGQBELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 120

DB 69 SSAGIMVAASPYVWWSVTAGQBELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 128

QY 121 LIATNLLGQNTPAIVNAEYGEWMAQDAAMFCYAAATATATATLIPFEAPEMTSAGG 180

DB 129 LIATNLLGQNTPAIVNAEYGEWMAQDAAMFCYAAATATATATLIPFEAPEMTSAGG 188

QY 181 LLEQAAAABEASDTAAANQLMNNVPQALQLOAQPTQCTPSSKLGGLWKTVPSPHRSPI 240

DB 189 LLEQAAAABEASDTAAANQLMNNVPQALQLOAQPTQCTPSSKLGGLWKTVPSPHRSPI 248

QY 241 MVSNNHMTNSGVSWTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300

DB 249 MVSNNHMTNSGVSWTNTLSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 308

QY 301 LGGVVAANLGRAASVGSLSVPCAWAAANQAVTPAARALPLTSLTSAARGPQOMLGGLPV 360
 DB 309 LGGVVAANLGRAASVGSLSVPCAWAAANQAVTPAARALPLTSLTSAARGPQOMLGGLPV 368
 QY 361 GQMGARAGGGLSGVLRVPPRPYPVMPHSPAAG 391
 DB 369 GQMGARAGGGLSGVLRVPPRPYPVMPHSPAAG 399

Search completed: June 30, 2004, 16:48:47
 Job time : 41.2057 secs

Sequence 16, Appl
Sequence 12, Appl
Sequence 142, App
Sequence 142, App
Sequence 131, App
Sequence 15, Appl
Sequence 15, Appl
Sequence 57, Appl
Sequence 92, Appl
Sequence 114, App
Sequence 109, App
Sequence 114, App
Sequence 109, App
Sequence 114, App
Sequence 2, Appl
Sequence 2, Appl
Sequence 2, Appl

US-09-287-849-16
US-09-287-849-12
US-09-073-009-142
US-09-073-010-142
US-09-477-135A-131
US-09-073-009-15
US-09-073-010-15
US-08-311-731A-57
US-08-311-731A-208
US-09-050-739-92
US-08-818-112-114
US-08-818-111-109
US-09-056-556-114
US-09-072-596-109
US-09-072-367-114
US-08-095-734-2
US-08-444-623-2
US-08-471-869-2

28 604 31.0 710 4
29 604 31.0 856 4
30 603 30.9 423 4
31 603 30.9 423 4
32 424.5 21.8 943 4
33 424 21.8 141 4
34 424 21.8 141 4
35 381.5 19.6 204 4
36 377.5 19.4 208 4
37 317 16.3 371 4
38 314 16.1 368 4
39 314 16.1 368 4
40 314 16.1 368 4
41 314 16.1 368 4
42 314 16.1 368 4
43 261 13.4 1271 1
44 261 13.4 1271 2
45 261 13.4 1271 3

ALIGNMENTS

RESULT 1
US-08-818-112-107
; Sequence 107, Application US/08818112
; Patent No. 6290369
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-VAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-107

Query Match 100.0%; Score 1949; DB 3; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:44:57 ; Search time 11.8942 Seconds
(without alignments)
1697.113 Million cell updates/sec

Title: US-09-597-796C-8
Perfect score: 1949
Sequence: 1 MVDFGALPEINSARMYAGP.....SGVLVPPRPVYMPHSPHAG 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/prodata/2/iaa/6C_COMB.pep.*
- 6: /cgn2_6/prodata/2/iaa/6D_COMB.pep.*
- 7: /cgn2_6/prodata/2/iaa/6E_COMB.pep.*
- 8: /cgn2_6/prodata/2/iaa/6F_COMB.pep.*
- 9: /cgn2_6/prodata/2/iaa/6G_COMB.pep.*
- 10: /cgn2_6/prodata/2/iaa/6H_COMB.pep.*
- 11: /cgn2_6/prodata/2/iaa/6I_COMB.pep.*
- 12: /cgn2_6/prodata/2/iaa/6J_COMB.pep.*
- 13: /cgn2_6/prodata/2/iaa/6K_COMB.pep.*
- 14: /cgn2_6/prodata/2/iaa/6L_COMB.pep.*
- 15: /cgn2_6/prodata/2/iaa/6M_COMB.pep.*
- 16: /cgn2_6/prodata/2/iaa/6N_COMB.pep.*
- 17: /cgn2_6/prodata/2/iaa/6O_COMB.pep.*
- 18: /cgn2_6/prodata/2/iaa/6P_COMB.pep.*
- 19: /cgn2_6/prodata/2/iaa/6Q_COMB.pep.*
- 20: /cgn2_6/prodata/2/iaa/6R_COMB.pep.*
- 21: /cgn2_6/prodata/2/iaa/6S_COMB.pep.*
- 22: /cgn2_6/prodata/2/iaa/6T_COMB.pep.*
- 23: /cgn2_6/prodata/2/iaa/6U_COMB.pep.*
- 24: /cgn2_6/prodata/2/iaa/6V_COMB.pep.*
- 25: /cgn2_6/prodata/2/iaa/6W_COMB.pep.*
- 26: /cgn2_6/prodata/2/iaa/6X_COMB.pep.*
- 27: /cgn2_6/prodata/2/iaa/6Y_COMB.pep.*
- 28: /cgn2_6/prodata/2/iaa/6Z_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1949	100.0	391	3	US-08-818-112-107
2	1949	100.0	391	4	US-08-818-111-102
3	1949	100.0	391	4	US-09-056-556-107
4	1949	100.0	391	4	US-09-072-596-102
5	1949	100.0	391	4	US-09-072-967-107
6	1949	100.0	596	4	US-09-287-849-26
7	1949	100.0	600	4	US-09-287-849-22
8	1944	99.7	729	4	US-09-223-040-2
9	1944	99.7	729	4	US-09-287-849-2
10	1852.5	84.8	396	3	US-08-818-112-111
11	1852.5	84.8	396	4	US-08-818-111-106
12	1852.5	84.8	396	4	US-09-056-556-111
13	1852.5	84.8	396	4	US-09-072-596-106
14	1852.5	84.8	396	4	US-09-072-967-111
15	1486.5	76.3	359	3	US-08-818-112-109
16	1486.5	76.3	359	4	US-08-818-111-104
17	1486.5	76.3	359	4	US-09-056-556-109
18	1486.5	76.3	359	4	US-09-072-596-104
19	1486.5	76.3	359	4	US-09-072-367-109
20	1187.5	60.9	358	4	US-09-287-849-8
21	1187.5	60.9	263	3	US-08-818-112-91
22	1187.5	60.9	263	4	US-08-818-111-92
23	1187.5	60.9	263	4	US-09-056-556-91
24	1187.5	60.9	263	4	US-09-072-596-92
25	1187.5	60.9	263	4	US-09-072-967-91
26	766.5	39.3	400	4	US-09-073-009-126
27	766.5	39.3	400	4	US-09-073-010-126

QY 1 MYDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
DB 1 MYDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
QY 61 SSAGLWVAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPFVIAENRAELMI 120
DB 61 SSAGLWVAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPFVIAENRAELMI 120
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEEAPEMTSAGG 180
DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEEAPEMTSAGG 180
QY 181 LLEQAAAEEASDTAAANQNMNVPAALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAAEEASDTAAANQNMNVPAALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSG 300
DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSG 300
QY 301 LGGGVAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
DB 301 LGGGVAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
QY 361 GQMGARAGGSLGVLRVPPRYVMPHSPAAG 391
DB 361 GQMGARAGGSLGVLRVPPRYVMPHSPAAG 391

RESULT 2

US-08-818-111-102
; Sequence 102: Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond S.
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.41706
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-111-102

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
DB 1 MYDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIG 60
QY 61 SSAGLWVAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPFVIAENRAELMI 120
DB 61 SSAGLWVAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPFVIAENRAELMI 120
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEEAPEMTSAGG 180
DB 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEEAPEMTSAGG 180
QY 181 LLEQAAAEEASDTAAANQNMNVPAALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAAEEASDTAAANQNMNVPAALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSG 300
DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSG 300
QY 301 LGGGVAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
DB 301 LGGGVAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 360
QY 361 GQMGARAGGSLGVLRVPPRYVMPHSPAAG 391
DB 361 GQMGARAGGSLGVLRVPPRYVMPHSPAAG 391

RESULT 3

US-09-056-556-107
; Sequence 107: Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
TREATME

US-09-056-556-107

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 1 MYDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 61 SSAGLMAAASPYVAMSVTAGOAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMAAASPYVAMSVTAGOAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLGONTTAAANQNMNVPOALQOLAQTQGTTPSSKLGGLMKTVPSPHSPI 240
DB 121 LIATNLGONTTAAANQNMNVPOALQOLAQTQGTTPSSKLGGLMKTVPSPHSPI 240
QY 181 LLEQAAVEEASDTAAANQNMNVPOALQOLAQTQGTTPSSKLGGLMKTVPSPHSPI 300
DB 181 LLEQAAVEEASDTAAANQNMNVPOALQOLAQTQGTTPSSKLGGLMKTVPSPHSPI 300
QY 241 MYSMANNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAARALPLTSLTSAERGPQMLGGLPV 360
DB 241 MYSMANNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAARALPLTSLTSAERGPQMLGGLPV 360
QY 301 LGGGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
DB 301 LGGGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
QY 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391
DB 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391

RESULT 4

US-09-072-596-102
Sequence 102, Application US/09072596
Patent No. 6458366
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 350
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/09/072,596
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-596-102

Query Match 100.0%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 2.9e-154;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
DB 1 MYDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 61 SSAGLMAAASPYVAMSVTAGOAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMAAASPYVAMSVTAGOAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLGONTTAAANQNMNVPOALQOLAQTQGTTPSSKLGGLMKTVPSPHSPI 240
DB 121 LIATNLGONTTAAANQNMNVPOALQOLAQTQGTTPSSKLGGLMKTVPSPHSPI 240
QY 181 LLEQAAVEEASDTAAANQNMNVPOALQOLAQTQGTTPSSKLGGLMKTVPSPHSPI 300
DB 181 LLEQAAVEEASDTAAANQNMNVPOALQOLAQTQGTTPSSKLGGLMKTVPSPHSPI 300
QY 241 MYSMANNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAARALPLTSLTSAERGPQMLGGLPV 360
DB 241 MYSMANNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAARALPLTSLTSAERGPQMLGGLPV 360
QY 301 LGGGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
DB 301 LGGGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALPLTSLTSAERGPQMLGGLPV 360
QY 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391
DB 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391

RESULT 5

US-09-072-967-107
Sequence 107, Application US/09072967
Patent No. 6592877
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/09/072,967

; FILING DATE: 05-MAY-1998
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.411C9
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 107:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 391 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; US-09-072-967-107

Query Match 100.0%; Score 1949; DB 4; Length 391;
 Best Local Similarity 100.0%; Pred. No. 2.9e-154;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGWSIG 60
 Db 1 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGWSIG 60
 QY 61 SSAGLWVAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 Db 61 SSAGLWVAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 QY 121 LIATNLLGONTPAIYVNEAEYGEWMAQDAAMFYAAATATATATLPPFEAPEMTSAGG 180
 Db 121 LIATNLLGONTPAIYVNEAEYGEWMAQDAAMFYAAATATATATLPPFEAPEMTSAGG 180
 QY 181 LLEQAAVEEASDSTAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
 Db 181 LLEQAAVEEASDSTAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
 QY 241 MVSMAHHMSTNSGVSMNTLSSMLKGFAPAAAAQVQTAQNGVRAMSSLGSSLGSSG 300
 Db 241 MVSMAHHMSTNSGVSMNTLSSMLKGFAPAAAAQVQTAQNGVRAMSSLGSSLGSSG 300
 QY 301 LGGGVAANLGRAASVGSLSVQAAWAAANQVTPAARALPLTSLTSAABERGQMLGGLPV 360
 Db 301 LGGGVAANLGRAASVGSLSVQAAWAAANQVTPAARALPLTSLTSAABERGQMLGGLPV 360
 QY 361 GQMGARAGGGLSGVLRVPPRPVPHSPAAG 391
 Db 361 GQMGARAGGGLSGVLRVPPRPVPHSPAAG 391

RESULT 6

US-09-287-849-26
 ; Sequence 26, Application US/09287849
 ; Patent No. 6627198

GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
 ; FILE REFERENCE: 014058-009020US
 ; CURRENT APPLICATION NUMBER: US/09/287,849
 ; CURRENT FILING DATE: 1999-04-07
 ; PRIOR APPLICATION NUMBER: US 08/818,112
 ; PRIOR FILING DATE: 1997-03-13
 ; PRIOR APPLICATION NUMBER: US 08/942,578
 ; PRIOR FILING DATE: 1997-10-01
 ; PRIOR APPLICATION NUMBER: US 09/025,197
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 09/056,556

; PRIOR FILING DATE: 1998-04-07
 ; PRIOR APPLICATION NUMBER: US 09/223,040
 ; PRIOR FILING DATE: 1998-12-30
 ; NUMBER OF SEQ ID NOS: 46
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO: 26
 ; LENGTH: 596
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
 ; US-09-287-849-26

Query Match 100.0%; Score 1949; DB 4; Length 596;
 Best Local Similarity 100.0%; Pred. No. 5.2e-154;
 Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGWSIG 60
 Db 9 MVDGALPPEINSARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGWSIG 68
 QY 61 SSAGLWVAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
 Db 69 SSAGLWVAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
 QY 121 LIATNLLGONTPAIYVNEAEYGEWMAQDAAMFYAAATATATATLPPFEAPEMTSAGG 180
 Db 129 LIATNLLGONTPAIYVNEAEYGEWMAQDAAMFYAAATATATATLPPFEAPEMTSAGG 188
 QY 181 LLEQAAVEEASDSTAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240
 Db 189 LLEQAAVEEASDSTAANQLMNNVPOALQQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 248
 QY 241 MVSMAHHMSTNSGVSMNTLSSMLKGFAPAAAAQVQTAQNGVRAMSSLGSSLGSSG 300
 Db 249 MVSMAHHMSTNSGVSMNTLSSMLKGFAPAAAAQVQTAQNGVRAMSSLGSSLGSSG 308
 QY 301 LGGGVAANLGRAASVGSLSVQAAWAAANQVTPAARALPLTSLTSAABERGQMLGGLPV 360
 Db 309 LGGGVAANLGRAASVGSLSVQAAWAAANQVTPAARALPLTSLTSAABERGQMLGGLPV 368
 QY 361 GQMGARAGGGLSGVLRVPPRPVPHSPAAG 391
 Db 369 GQMGARAGGGLSGVLRVPPRPVPHSPAAG 399

RESULT 7

US-09-287-849-22
 ; Sequence 22, Application US/09287849
 ; Patent No. 6627198

GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
 ; FILE REFERENCE: 014058-009020US
 ; CURRENT APPLICATION NUMBER: US/09/287,849
 ; CURRENT FILING DATE: 1999-04-07
 ; PRIOR APPLICATION NUMBER: US 08/818,112
 ; PRIOR FILING DATE: 1997-03-13
 ; PRIOR APPLICATION NUMBER: US 08/942,578
 ; PRIOR FILING DATE: 1997-10-01
 ; PRIOR APPLICATION NUMBER: US 09/025,197
 ; PRIOR FILING DATE: 1998-02-18
 ; PRIOR APPLICATION NUMBER: US 09/056,556
 ; PRIOR FILING DATE: 1998-04-07
 ; PRIOR APPLICATION NUMBER: US 09/223,040
 ; PRIOR FILING DATE: 1998-12-30
 ; NUMBER OF SEQ ID NOS: 46

QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPBMTSAGG 180
DB 262 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPBMTSAGG 321
QY 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 322 LLEQAAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 381
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 300
DB 382 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 441
QY 301 LGGGVAANLGRAASVGSLSVFPQAAANQAVTPAARALPLTSLTSAERGPQMLGLPV 360
DB 442 LGGGVAANLGRAASVGSLSVFPQAAANQAVTPAARALPLTSLTSAERGPQMLGLPV 501
QY 361 GOMGARAGGGLSGVLRVPRPYVMPHSPAAG 391
DB 502 GOMGARAGGGLSGVLRVPRPYVMPHSPAAG 532

RESULT 10

US-08-818-112-111
; Sequence 111, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

US-08-818-112-111

Query Match 84.8%; Score 1652.5; DB 3; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.4e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTGSWIG 60

DB 1 VVDFGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTGSWIG 60
QY 61 SSAGLVAAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPRVIAENAEELMI 120
DB 61 SSAGLVAAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPRVIAENAEELMI 120
QY 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPBMTSAGG 180
DB 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPBMTSAGG 180
QY 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
DB 181 LLEQAAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 240
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 296
DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSG 296
QY 297 GSSGLGGGVAANLGRAASVGSLSVFPQAAANQAVTPAARALPLTSLTSAERGPQMLG 356
DB 300 GSSGLGGGVAANLGRAASVGSLSVFPQAAANQAVTPAARALPLTSLTSAERGPQMLG 359
QY 357 GLPVGOMGARAG--GGLSGVLRVPRPYVMPHSPAAG 391
DB 360 GLPLGOLTNSGGGFGVSNALMPPRVYVMPRVPAAAG 396

RESULT 11

US-08-818-111-106
; Sequence 106, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

US-08-818-111-106

US-08-818-111-106

Query Match	84.8%	Score 1652.5	DB 4	Length 396
Best Local Similarity	84.9%	Pred. No. 1.4e-129		
Matches 337	Conservative 19	Mismatches 34	Indels 7	Gaps 3
QY	1	MYDFCALPPEINSAARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTGVSGWIG	60	
DB	1	VVDFCALPPEINSAARMYAGPGSASLVAAAQWMDSVASDLFSAASAFQSVVWGLTGVSGWIG	60	
QY	61	SSAGLMVAAAAPYVAMWSVTAGQELTAQAQVRVAAAAAYETAYGLTVPPPVIAENRAELMI	120	
DB	61	SSAGLMVAAAAPYVAMWSVTAGQELTAQAQVRVAAAAAYETAYGLTVPPPVIAENRAELMI	120	
QY	121	LIATNLLGQNTPATVNEAEYGEWMAODAAAAMFGYAAATATATATLPPPEEAPEMTSAGG	180	
DB	121	LIATNLLGQNTPATVNEAEYGEWMAODAAAAMFGYAAATATATLPPPEEAPEMTSAGG	180	
QY	181	LLEQAAAVEEASDTAAANQLMNNYPQALQLOAQPTQGTTPSSKLGGLWKTVSEHRSPTSN	240	
DB	181	LLEQAAVEEAEIDTAAANQLMNNYPQALQLOAQPTKSIWPFQDLSLWKAISPHLSPLSN	240	
QY	241	MYSMANNHWSMTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVYRMSS---	296	
DB	241	IVSMLNHNHWSMTNSGVSMNASTLHSMKGFAP-AAAQAVETAAQNGVQVAMSSLGSGQJSSL	299	
QY	297	GSSGLGGVVAANLGRASAVGSLSPVQAWAANAQVTPAAPALPLTSLTSAABRGPQOMLG	356	
DB	300	GSSGLGAGVAANLGEAAAVGSLSPVQAWAANAQVTPAARALPLTSLTSAAQTPAGHMIG	359	
QY	357	GLPVGOMGARAG--GGLSGVLRVPRPYPVWPHSPAG	391	
DB	360	GLPLGOLTNISGGFGGVSNALMPFRAYVWPRVPAAG	396	

TREATMENT

RESULT 12

US-09-056-556-111

; Sequence 111, Application US/09056556

; Patent No. 6350456

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND

; NUMBER OF SEQUENCES: 241

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: SEED AND BERRY LLP

; STREET: 6300 Columbia Center, 701 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98104-7092

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/056,556

; FILING DATE: 07-APR-1998

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Maki, David J.

; REGISTRATION NUMBER: 31,392

; REFERENCE/DOCKET NUMBER: 210121.457

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 622-4900

; TELEFAX: (206) 682-6031

; INFORMATION FOR SEQ ID NO: 111:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 396 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

US-09-056-556-111

RESULT 13
US-09-072-596-i06
; Sequence 106, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 106:

SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-596-106

Query Match 84.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.4e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 1 MVDGALPPEINSARMYAGPGSASIVAAAKWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
DB 1 VVDFGALPPEINSARMYAGPGSASIVAAAKWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
QY 61 SSAGLMVAAASPYVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
DB 61 SSAGLMVAAASPYVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
QY 121 LIATNLLGONTFAIVNEAEYGEWMAQDAAMFGVAAATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLLGONTFAIVNEAEYGEWMAQDAAMFGVAAATATATALLPFEDAPLITNPGG 180
QY 181 LLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGTTPSSKGLGKLVKTVSPHRSPISN 240
DB 181 LLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGTTPSSKGLGKLVKTVSPHRSPISN 240
QY 241 MVSMMNNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQAONGVAMSS ---LGSSL 296
DB 241 IVSMLNNHVSMTNSGVSMASTLHSLMKGFAP-AAAQAVETAQAONGVAMSSLGSLGSSL 299
QY 297 GSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAARPGOMLG 356
DB 300 GSSGLGAGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAQAQAPGHMLG 359
QY 357 GLPVQMGARAG--GGLSGVLVPRPYVMPHSPAAG 391
DB 360 GLPLQLTNSGGGFGVSNALRMPPRAYVMPRVPAAG 396

RESULT 14

US-09-072-967-111
Sequence 111, Application US/09072967
Patent No. 6592877

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Wedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/072,967

FILING DATE: 05-MAY-1998

CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:

NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 111:

SEQUENCE CHARACTERISTICS:

LENGTH: 396 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

US-09-072-967-111

Query Match 84.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 1.4e-129;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 1 MVDGALPPEINSARMYAGPGSASIVAAAKWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
DB 1 VVDFGALPPEINSARMYAGPGSASIVAAAKWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
QY 61 SSAGLMVAAASPYVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
DB 61 SSAGLMVAAASPYVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
QY 121 LIATNLLGONTFAIVNEAEYGEWMAQDAAMFGVAAATATATATLLPFEAPEMTSAGG 180
DB 121 LIATNLLGONTFAIVNEAEYGEWMAQDAAMFGVAAATATATALLPFEDAPLITNPGG 180
QY 181 LLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGTTPSSKGLGKLVKTVSPHRSPISN 240
DB 181 LLEQAAVEEASDTAAANQLMNNVPOALQQLAQTQGTTPSSKGLGKLVKTVSPHRSPISN 240
QY 241 MVSMMNNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQAONGVAMSS ---LGSSL 296
DB 241 IVSMLNNHVSMTNSGVSMASTLHSLMKGFAP-AAAQAVETAQAONGVAMSSLGSLGSSL 299
QY 297 GSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAARPGOMLG 356
DB 300 GSSGLGAGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAQAQAPGHMLG 359
QY 357 GLPVQMGARAG--GGLSGVLVPRPYVMPHSPAAG 391
DB 360 GLPLQLTNSGGGFGVSNALRMPPRAYVMPRVPAAG 396

RESULT 15

US-08-818-112-109

Sequence 109, Application US/08818112

Patent No. 6290969

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.

APPLICANT: Dillon, Davin C.

APPLICANT: Campos-Neto, Antonio

APPLICANT: Houghton, Raymond

APPLICANT: Wedvick, Thomas S.

APPLICANT: Twardzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY

TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 153

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk


```

;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 109:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 359 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-818-112-109

Query Match 76.3%; Score 1486.5; DB 3; Length 359;
Best Local Similarity 84.2%; Pred. No. 7.9e-116;
Matches 303; Conservative 16; Mismatches 36; Indels 5; Gaps 2;

Qy 1 MVDFGALPPPEINARMYAGSGASLVAAAKMWDVSDLFSAASAFQSVVWGLTVGSGWIG 60
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 1 VVDFGALPPPEINARMYAGSGASLVAAAKMWDVSDLFSAASAFQSVVWGLTVGSGWIG 60
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 61 SSAGLMVAAASPVVAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 61 SSAGLMVAAASPVVAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMT 120
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 121 LIATNLIGQNTPAIAVNEAYGEMWQDAAMFGYAAATATATATATLLPPEAPEMTSAGG 180
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 121 LTATNLIGQNTPAIEANQAAYSQWQDAEAMYGAAATATATATEALLPFEDAPLIITNPGG 180
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 181 LLEQAAVAVEASDTAAANQLMNNVPOALQOLAQPTGCTTPSSKLGGLWKTVSPHRSPISN 240
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 181 LLEQAAVAVEAIDTAAANQLMNNVPOALQOLAQPTGCTTPSSKLGGLWKTAVSPHLSPLSN 240
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 241 MVSMAHHMNSMTNSGVSMNTTSLSMKGFAPAAAQAQVOTAAQNGVRAVSS----LGSSL 296
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 241 VSSIANNHMSMMGTGVSMTNTLHSMKGLAP-AAAQAVETAENGVMWSSSLGSLGSSL 299
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 297 GSSGLGCVAAANLGRAASVGSLSVPOAWAANQAVTPAARALPITSLSAERGPQOMLG 356
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Qy 300 GSSGLGAGVAAANLGRAASVGSLSVPPAWAANQAVTPAARALPITSLSAATAPGHMLG 359
Db :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Search completed: June 30, 2004, 16:55:19
Job time : 12.8942 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:52:58 ; Search time 30.4893 Seconds
(without alignments)
3625.462 Million cell updates/sec

Title: US-09-597-796C-8

Perfect score: 1949
Sequence: 1 MYDFGALPPEINSAARMYAGP.....SGVLRVPRPYMPHSPAG 391

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1166195 seqs, 282705291 residues

Total number of hits satisfying chosen parameters: 1166195

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PTCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PTCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1949	100.0	391	12	US-09-886-349A-14
2	1949	100.0	391	14	US-10-193-002-102
3	1949	100.0	391	14	US-10-084-843-107
4	1949	100.0	391	14	US-10-098-732A-14
5	1949	100.0	596	9	US-09-287-849-26
6	1949	100.0	596	12	US-09-886-349A-20
7	1949	100.0	596	14	US-10-359-460-26
8	1949	100.0	596	14	US-10-098-732A-20
9	1949	100.0	600	9	US-09-287-849-22
10	1949	100.0	600	14	US-10-359-460-22
11	1949	100.0	723	15	US-10-369-983-2
12	1949	100.0	729	12	US-09-886-349A-18
13	1949	100.0	729	14	US-10-098-732A-18
14	1949	100.0	729	15	US-10-369-983-21
15	1949	100.0	729	15	US-10-369-983-22

16	1949	100.0	813	15	US-10-369-983-15	Sequence 15, Appl
17	1949	100.0	825	15	US-10-369-983-14	Sequence 14, Appl
18	1949	100.0	875	15	US-10-369-983-13	Sequence 13, Appl
19	1949	100.0	930	14	US-10-098-732A-65	Sequence 65, Appl
20	1949	100.0	930	15	US-10-369-983-12	Sequence 12, Appl
21	1949	100.0	1010	15	US-10-369-983-4	Sequence 4, Appl
22	1949	100.0	1016	15	US-10-369-983-18	Sequence 18, Appl
23	1949	100.0	1022	15	US-10-369-983-17	Sequence 17, Appl
24	1949	100.0	1154	15	US-10-369-983-16	Sequence 16, Appl
25	1944	99.7	729	9	US-09-287-849-2	Sequence 2, Appl
26	1944	99.7	729	12	US-09-886-349A-16	Sequence 16, Appl
27	1944	99.7	729	14	US-10-359-460-2	Sequence 2, Appl
28	1944	99.7	729	14	US-10-098-732A-16	Sequence 16, Appl
29	1944	99.7	729	15	US-10-359-459-2	Sequence 2, Appl
30	1934	99.2	391	12	US-09-872-186-8	Sequence 8, Appl
31	1652.5	84.8	396	14	US-10-193-002-106	Sequence 106, App
32	1652.5	84.8	396	14	US-10-084-843-111	Sequence 111, App
33	1583	81.2	393	12	US-10-282-122A-62455	Sequence 62455, A
34	1583	81.2	393	12	US-10-282-122A-64892	Sequence 64892, A
35	1486.5	76.3	359	14	US-10-193-002-104	Sequence 104, App
36	1486.5	76.3	359	14	US-10-084-843-109	Sequence 109, App
37	1187.5	60.9	358	9	US-09-287-849-8	Sequence 8, Appl
38	1187.5	60.9	358	14	US-10-359-460-8	Sequence 8, Appl
39	1187	60.9	263	12	US-09-886-349A-12	Sequence 12, Appl
40	1187	60.9	263	14	US-10-193-002-92	Sequence 92, Appl
41	1187	60.9	263	14	US-10-084-843-91	Sequence 91, Appl
42	1187	60.9	263	14	US-10-098-732A-12	Sequence 12, Appl
43	766.5	39.3	400	9	US-09-073-009-126	Sequence 126, App
44	766.5	39.3	400	9	US-09-793-306-126	Sequence 126, App
45	742.5	38.1	405	12	US-10-282-122A-62027	Sequence 62027, A

ALIGNMENTS

RESULT 1
US-09-886-349A-14
; Sequence 14, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTE39 (TBH9FL)
US-09-886-349A-14

Query Match 100.0%; Score 1949; DB 12; Length 391;
Best Local Similarity 100.0%; Pred. No. 1e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	MYDFGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
DB	1	MYDFGALPPEINSAARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSWIG 60
QY	61	SSAGLWVAASPYVAVMSVTAGQABLTAAQVRVAAAAYETAYGLTVPFPVIAENRAELMI 120
DB	61	SSAGLWVAASPYVAVMSVTAGQABLTAAQVRVAAAAYETAYGLTVPFPVIAENRAELMI 120

REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 107:
US-10-084-843-107

Query Match
Best Local Similarity 100.0%; Score 1949; DB 14; Length 391;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
Db 1 MVDGALPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
QY 61 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
Db 61 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
QY 121 LIATNLIGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180
Db 121 LIATNLIGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180
QY 181 LLEQAAVVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPI 240
Db 181 LLEQAAVVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPI 240
QY 241 MVSMAHHMSTNGSVMTNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSSG 300
Db 241 MVSMAHHMSTNGSVMTNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSSG 300
QY 301 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
Db 301 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
QY 361 GQMGARAGGSLGVLVPRPPYVMPHSPAAG 391
Db 361 GQMGARAGGSLGVLVPRPPYVMPHSPAAG 391

RESULT 4
US-10-098-732A-14
Sequence 14, Application US/10098732A
Publication No. US20030175294A1
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Brannon, Mark
APPLICANT: Guderian, Jeffrey
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
TITLE OF INVENTION: Leishmania Antigen
FILE REFERENCE: 014058-012010US
CURRENT APPLICATION NUMBER: US/10/098,732A
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: US 60/275,837
PRIOR FILING DATE: 2001-03-13
NUMBER OF SEQ ID NOS: 80
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 14
LENGTH: 391
TYPE: PRT
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: MTB39 full length (TBH9FL)
US-10-098-732A-14

Query Match
100.0%; Score 1949; DB 14; Length 391;

Best Local Similarity 100.0%; Pred. No. 1e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
Db 1 MVDGALPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60
QY 61 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
Db 61 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120
QY 121 LIATNLIGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180
Db 121 LIATNLIGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 180
QY 181 LLEQAAVVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPI 240
Db 181 LLEQAAVVEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPI 240
QY 241 MVSMAHHMSTNGSVMTNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSSG 300
Db 241 MVSMAHHMSTNGSVMTNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSSG 300
QY 301 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
Db 301 LGGVVAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
QY 361 GQMGARAGGSLGVLVPRPPYVMPHSPAAG 391
Db 361 GQMGARAGGSLGVLVPRPPYVMPHSPAAG 391

RESULT 5
US-09-287-849-26
Sequence 26, Application US/09287849
Patent No. US20020009459A1
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
TITLE OF INVENTION: and Their Uses
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/09/287,849
CURRENT FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 26
LENGTH: 596
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: bi-fusion
US-09-287-849-26

Query Match
Best Local Similarity 100.0%; Pred. No. 1.8e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSGWIG 60

QY 241 MVSANNHMTNSGVSMTNTLSSMLKGFAPAAAAQAVTAQAQNGVRAMSSSLGSSG 300
Db 249 MVSANNHMTNSGVSMTNTLSSMLKGFAPAAAAQAVTAQAQNGVRAMSSSLGSSG 308
QY 301 LGGVAAANLGRAASVGSLSVPOQAAWAAANQAVTPAARALPLTSLTSAARGPQOMLGLPLV 360
Db 309 LGGVAAANLGRAASVGSLSVPOQAAWAAANQAVTPAARALPLTSLTSAARGPQOMLGLPLV 368
QY 361 GOMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
Db 369 GOMGARAGGGLSGVLRVPPRPVMPHSPAAG 399

RESULT 8

US-10-098-732A-20
; Sequence 20, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TBH9-Ra35 (designated MTB59P)
US-10-098-732A-20

Query Match 100.0%; Score 1949; DB 14; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.8e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
Db 9 MVDGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 68
QY 61 SSAGLVAAASPVYAWMSVTAQAELTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
Db 69 SSAGLVAAASPVYAWMSVTAQAELTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 128
QY 121 LIATNLLGQNTPAIAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
Db 129 LIATNLLGQNTPAIAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
QY 181 LLEQAAAVEASDTAAANLNQNNVPOALQLOAQTOGTTPSSKLGGLWKTVPSPHSPISN 240
Db 189 LLEQAAAVEASDTAAANLNQNNVPOALQLOAQTOGTTPSSKLGGLWKTVPSPHSPISN 248
QY 241 MVSANNHMTNSGVSMTNTLSSMLKGFAPAAAAQAVTAQAQNGVRAMSSSLGSSG 300
Db 249 MVSANNHMTNSGVSMTNTLSSMLKGFAPAAAAQAVTAQAQNGVRAMSSSLGSSG 308
QY 301 LGGVAAANLGRAASVGSLSVPOQAAWAAANQAVTPAARALPLTSLTSAARGPQOMLGLPLV 360
Db 309 LGGVAAANLGRAASVGSLSVPOQAAWAAANQAVTPAARALPLTSLTSAARGPQOMLGLPLV 368
QY 361 GOMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
Db 369 GOMGARAGGGLSGVLRVPPRPVMPHSPAAG 399

RESULT 10

US-10-359-460-22
; Sequence 22, Application US/10359460

RESULT 9
US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

Query Match 100.0%; Score 1949; DB 9; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.8e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 60
Db 9 MVDGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 68
QY 61 SSAGLVAAASPVYAWMSVTAQAELTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
Db 69 SSAGLVAAASPVYAWMSVTAQAELTAAQVRVAAAAYETAYGLTVPPIAENRAELMI 128
QY 121 LIATNLLGQNTPAIAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 180
Db 129 LIATNLLGQNTPAIAVNEAEYGENWAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
QY 181 LLEQAAAVEASDTAAANLNQNNVPOALQLOAQTOGTTPSSKLGGLWKTVPSPHSPISN 240
Db 189 LLEQAAAVEASDTAAANLNQNNVPOALQLOAQTOGTTPSSKLGGLWKTVPSPHSPISN 248
QY 241 MVSANNHMTNSGVSMTNTLSSMLKGFAPAAAAQAVTAQAQNGVRAMSSSLGSSG 300
Db 249 MVSANNHMTNSGVSMTNTLSSMLKGFAPAAAAQAVTAQAQNGVRAMSSSLGSSG 308
QY 301 LGGVAAANLGRAASVGSLSVPOQAAWAAANQAVTPAARALPLTSLTSAARGPQOMLGLPLV 360
Db 309 LGGVAAANLGRAASVGSLSVPOQAAWAAANQAVTPAARALPLTSLTSAARGPQOMLGLPLV 368
QY 361 GOMGARAGGGLSGVLRVPPRPVMPHSPAAG 391
Db 369 GOMGARAGGGLSGVLRVPPRPVMPHSPAAG 399

```

; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO. 2
; LENGTH: 723
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mutated
; OTHER INFORMATION: MTB32-MTB33F fusion protein (MTB32MutSA)
US-10-369-983-2

Query Match 100.0%; Score 1949; DB 15; Length 723;
Best Local Similarity 100.0%; Pred. No. 2.3e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDGALPPEINSRMVYAGPGSASLVAAQMWDSVASDLFSAASAFAQSVVWGLTVGSWIG 60
DB 333 MVDGALPPEINSARMYTAGPGSASLVAAQMWDSVASDLFSAASAFAQSVVWGLTVGSWIG 392
QY 61 SSAGLWMAAASPYYAMVSVTAGQAEITAAQVRAAAAAYETAYGLTVPPPVIAENRAELMI 120
DB 393 SSAGLWMAAASPYYAMVSVTAGQAEITAAQVRAAAAAYETAYGLTVPPPVIAENRAELMI 452
QY 121 LIATNLIGQNTPAIVNEAEYEGEMWAQDAADAAAFGAAATATATATLPPFEAPEMTSAGG 180
DB 453 LIATNLIGQNTPAIVNEAEYEGEMWAQDAADAAAFGAAATATATATLPPFEAPEMTSAGG 512
QY 181 LLEQAAAVEEASDTAAANOLMNNVPOALQOLAQPTQGTTPSKLGLWKTVSPHRSPISN 240
DB 513 LLEQAAAVEEASDTAAANOLMNNVPOALQOLAQPTQGTTPSKLGLWKTVSPHRSPISN 572
QY 241 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLSGSSLSGSSG 300
DB 573 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLSGSSLSGSSG 632
QY 301 LGGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABRGPGQMLGLPLV 360
DB 633 LGGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAABRGPGQMLGLPLV 692
QY 361 GOMGARAGGSLGVLVPPRPYVMPHSPAAG 391
DB 693 GOMGARAGGSLGVLVPPRPYVMPHSPAAG 723

RESULT 12
US-09-886-349A-18
; Sequence 18, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1

```


QY	1	WVDFGALPPPEINSA	RMYPAGPGSASLVAA	AAQWMDSVASDLFSA	ASAFQSVVWGLTVGS	WIG	60
Db	142	WVDFGALPPPEINSA	RMYPAGPGSASLVAA	AAQWMDSVASDLFSA	ASAFQSVVWGLTVGS	WIG	201
QY	61	SSAGLVAAASPYVAM	SVTAQAEILTAQVR	VAAAAAYETAYGLT	VPVPIAENRAELMI	120	
Db	202	SSAGLVAAASPYVAM	SVTAQAEILTAQVR	VAAAAAYETAYGLT	VPVPIAENRAELMI	261	
QY	121	LIATNLLGONTPAI	AVNEAEYGEWMAQDA	AAAFGYAAATATATA	TATATLLPPEAP	EMTSAGG	180
Db	262	LIATNLLGONTPAI	AVNEAEYGEWMAQDA	AAAFGYAAATATATA	TATATLLPPEAP	EMTSAGG	321
QY	181	LLEQAAVEEASDTA	AAANQLMNNVPOAL	QOLAQPTQGTTPSSK	LGLWKTVP	SPHRSPISN	240
Db	322	LLEQAAVEEASDTA	AAANQLMNNVPOAL	QOLAQPTQGTTPSSK	LGLWKTVP	SPHRSPISN	381
QY	241	MYSMANNHSMTNS	GYVSMNTLSSMLKGP	APAAAAQAVOTAAQNG	VRAMSSLGSSG	300	
Db	382	MYSMANNHSMTNS	GYVSMNTLSSMLKGP	APAAAAQAVOTAAQNG	VRAMSSLGSSG	441	
QY	301	LGCGVAANI	GRAASVCSLVPQAW	AAANQAVTPAAEAL	PLTSLTSA	EAERPGQMLG	360
Db	442	LGCGVAANI	GRAASVCSLVPQAW	AAANQAVTPAAEAL	PLTSLTSA	EAERPGQMLG	501
QY	361	GQMGARAGGGLSV	LVRPPRPYPVMPH	SPAAG	391		
Db	502	GQMGARAGGGLSV	LVRPPRPYPVMPH	SPAAG	532		
RESULT 14							
US-10-369-983-21							
; Sequence 21, Application US/10369983							
; Publication No. US20030235593A1							
; GENERAL INFORMATION:							
; APPLICANT: Skeiky, Yasir							
; APPLICANT: Guderian, Jeff							
; APPLICANT: Reed, Steven							
; APPLICANT: Corixa Corporation							
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis							
; FILE REFERENCE: 014058-009081US							
; CURRENT APPLICATION NUMBER: US/10/369,983							
; CURRENT FILING DATE: 2003-02-18							
; PRIOR APPLICATION NUMBER: US 60/357,351							
; PRIOR FILING DATE: 2002-02-15							
; NUMBER OF SEQ ID NOS: 22							
; SOFTWARE: PatentIn Ver. 2.1							
; SEQ. ID. NO 21							
; LENGTH: 729							
; TYPE: PRT							
; ORGANISM: Artificial Sequence							
; FEATURE:							
; OTHER INFORMATION: Description of Artificial Sequence:MTB72F							
US-10-369-983-21							
Query Match 100.0%; Score 1949; DB 15; Length 729;							
Best Local Similarity 100.0%; Pred. No. 2.4e-145; Indels 0; Gaps 0;							
Matches 391; Conservative 0; Mismatches 0;							
QY	1	WVDFGALPPPEINSA	RMYPAGPGSASLVAA	AAQWMDSVASDLFSA	ASAFQSVVWGLTVGS	WIG	60
Db	142	WVDFGALPPPEINSA	RMYPAGPGSASLVAA	AAQWMDSVASDLFSA	ASAFQSVVWGLTVGS	WIG	201
QY	61	SSAGLVAAASPYVAM	SVTAQAEILTAQVR	VAAAAAYETAYGLT	VPVPIAENRAELMI	120	
Db	202	SSAGLVAAASPYVAM	SVTAQAEILTAQVR	VAAAAAYETAYGLT	VPVPIAENRAELMI	261	
QY	121	LIATNLLGONTPAI	AVNEAEYGEWMAQDA	AAAFGYAAATATATA	TATATLLPPEAP	EMTSAGG	180
Db	262	LIATNLLGONTPAI	AVNEAEYGEWMAQDA	AAAFGYAAATATATA	TATATLLPPEAP	EMTSAGG	321
QY	181	LLEQAAVEEASDTA	AAANQLMNNVPOAL	QOLAQPTQGTTPSSK	LGLWKTVP	SPHRSPISN	240
Db	322	LLEQAAVEEASDTA	AAANQLMNNVPOAL	QOLAQPTQGTTPSSK	LGLWKTVP	SPHRSPISN	381
QY	241	MYSMANNHSMTNS	GYVSMNTLSSMLKGP	APAAAAQAVOTAAQNG	VRAMSSLGSSG	300	
Db	382	MYSMANNHSMTNS	GYVSMNTLSSMLKGP	APAAAAQAVOTAAQNG	VRAMSSLGSSG	441	
QY	301	LGCGVAANI	GRAASVCSLVPQAW	AAANQAVTPAAEAL	PLTSLTSA	EAERPGQMLG	360
Db	442	LGCGVAANI	GRAASVCSLVPQAW	AAANQAVTPAAEAL	PLTSLTSA	EAERPGQMLG	501
QY	361	GQMGARAGGGLSV	LVRPPRPYPVMPH	SPAAG	391		
Db	502	GQMGARAGGGLSV	LVRPPRPYPVMPH	SPAAG	532		

```

; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72FmutSA
; OTHER INFORMATION: (Ra12-TbHP-Ra35MutSA)
US-09-886-349A-18

Query Match 100.0%; Score 1949; DB 12; Length 729;
Best Local Similarity 100.0%; Pred. No. 2.4e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSARVMYAGSGSASLVAAQAQMDVSDADLFSAAAFQSVVWGLTVGSWIG 60
Db 142 MVDFGALPPEINSARVMYAGSGSASLVAAQAQMDVSDADLFSAAAFQSVVWGLTVGSWIG 201
QY 61 SSAGLMAAASPVVAVMWSVTAGQAELTAAQVRVAAAAYETAYGLTVPPPPVIAENRAELMI 120
Db 202 SSAGLMAAASPVVAVMWSVTAGQAELTAAQVRVAAAAYETAYGLTVPPPPVIAENRAELMI 261
QY 121 LIATNLLGQNTPAIAVNEAEYGEWQAQDAAMFGYAAATATATATALLPPEAREPMTSAGG 180
Db 262 LIATNLLGQNTPAIAVNEAEYGEWQAQDAAMFGYAAATATATATALLPPEAREPMTSAGG 321
QY 181 LIEQAAAVEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGLLWKTIVSPHRSPISN 240
Db 322 LIEQAAAVEASDTAAANQLMNNVPQALQQLAQPTQGTTPSSKLGLLWKTIVSPHRSPISN 381
QY 241 MVSMAANHSMNTSGVSMNTLTLSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSLGSSG 300
Db 382 MVSMAANHSMNTSGVSMNTLTLSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSLGSSG 441
QY 301 LGGGVAANTLGRAASVGSLSVPQAWAANQAVTTPAARALPLTSLTSAERGFQMLGGLPV 360
Db 442 LGGGVAANTLGRAASVGSLSVPQAWAANQAVTTPAARALPLTSLTSAERGFQMLGGLPV 501
QY 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391
Db 502 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 532

RESULT 13
US-10-098-732A-18
; Sequence 18, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasaki
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098, 732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72FmutSA
; OTHER INFORMATION: (Ra12-TbH9-Ra35MutSA)
US-10-098-732A-18

Query Match 100.0%; Score 1949; DB 14; Length 729;
Best Local Similarity 100.0%; Pred. No. 2.4e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Search completed: June 30, 2004, 17:14:44
Job time : 31.4893 secs

```
QY 241 MVSNNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
Db 382 MVSNNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 441
QY 301 LGGGVAANLGRAASVGSLSVQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
Db 442 LGGGVAANLGRAASVGSLSVQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 501
QY 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391
Db 502 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 532
```

```
RESULT 15
US-10-369-983-22
; Sequence 22, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 22
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:mutated
; OTHER INFORMATION: MTB72F-mutSA (Mcb72f-mutSA)
US-10-369-983-22
```

```
Query Match 100.0%; Score 1949; DB 15; Length 729;
Best Local Similarity 100.0%; Pred. No. 2.4e-145;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDLSAASAFQSVVWGLTVGSWIG 60
Db 142 MVDFGALPPEINSARMYAGPGSASLVAAAQWDSVASDLSAASAFQSVVWGLTVGSWIG 201
QY 61 SSAGLMVAASPYVAVMSVTAGQAEATAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120
Db 202 SSAGLMVAASPYVAVMSVTAGQAEATAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 261
QY 121 LIATNLGQNTPAJAVNEAYGEMWAQDAAAFGYAAATATATATATLPPFEAPMTSAGG 180
Db 262 LIATNLGQNTPAJAVNEAYGEMWAQDAAAFGYAAATATATATLPPFEAPMTSAGG 321
QY 181 LLEQAAVEASDPAANQAMNNVPQALQLOAQTOGTTTPSSKLGGLKTKVSHRSPISN 240
Db 322 LLEQAAVEASDPAANQAMNNVPQALQLOAQTOGTTTPSSKLGGLKTKVSHRSPISN 381
QY 241 MVSNNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 300
Db 382 MVSNNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG 441
QY 301 LGGGVAANLGRAASVGSLSVQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 360
Db 442 LGGGVAANLGRAASVGSLSVQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 501
QY 361 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 391
Db 502 GQMGARAGGGLSGVLRVPPRYVMPHSPAAG 532
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:31 ; Search time 10.2189 Seconds
(without alignments)
3680.509 Million cell updates/sec

Title: US-09-597-796C-8

Perfect score: 1949
Sequence: 1 MVDGALPPEINARMYAGP.....SGVLVPRPRVMPHSPAAG 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

PIR 78: *
1: Pirl: *
2: Pirl: *
3: Pirl: *
4: Pirl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1949	100.0	391	2 B70608	probable PPE prote
2	1566.5	85.0	396	2 B70741	probable PPE prote
3	1583	81.2	393	2 C70568	probable PPE prote
4	779.5	40.0	393	2 G70929	probable PPE prote
5	751	38.5	409	2 A70932	probable PPE prote
6	737.5	37.8	403	2 B70931	probable PPE prote
7	736	37.8	423	2 H70931	probable PPE prote
8	731	37.5	421	2 H87056	PPE-family protein
9	705	36.2	408	2 G70925	probable PPE prote
10	702	36.0	391	2 B70625	probable PPE prote
11	702	36.0	413	2 F70560	probable PPE prote
12	688.5	35.3	463	2 C70931	probable PPE prote
13	682.5	35.0	468	2 B70932	probable PPE prote
14	675	34.6	380	2 A70646	probable PPE prote
15	668.5	34.3	394	2 G70881	probable PPE prote
16	667.5	34.2	385	2 H70903	probable PPE prote
17	635	32.6	350	2 H70929	probable PPE prote
18	633.5	32.5	365	2 E70929	probable PPE prote
19	615.5	31.6	402	2 A70882	probable PPE prote
20	603	30.9	423	2 C70582	probable PPE prote
21	597.5	30.7	391	2 D70922	probable PPE prote
22	589	30.2	406	2 E70875	probable PPE prote
23	588	30.2	394	2 A70504	probable PPE prote
24	575.5	29.5	391	2 A70663	probable PPE prote
25	495.5	25.4	3300	2 D70575	probable PPE prote
26	473	24.3	180	2 G70834	probable PPE prote
27	452.5	23.2	3716	2 E70969	probable PPE prote
28	448.5	23.0	580	2 G70570	probable PPE prote
29	444	22.8	346	2 H70874	probable PPE prote

ALIGNMENTS

RESULT 1

B70608

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: B70608

R: Cole, S.I.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feldwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: B70608

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-391 <COL>

A:Cross-references: GB:Z93777; GB:AL123456; NID:G3261726; PIDN:CAB07839.1; PID:e311073; I

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match	Score	DB	Length
Best Local Similarity	100.0%;	100.0%;	391;
Mismatches	0;	Mismatches	0;
Indels	0;	Gaps	0;
Matches	391;	Conservative	0;
QY	1	MVDGALPPEINARMYAGP	60
DB	1	MVDGALPPEINARMYAGP	60
QY	61	SSAGLMVAASPYVAMSVT	120
DB	61	SSAGLMVAASPYVAMSVT	120
QY	121	LIATNLLGQNTPAIAVNEA	180
DB	121	LIATNLLGQNTPAIAVNEA	180
QY	181	LLEQAAVEEASDTAAAN	240
DB	181	LLEQAAVEEASDTAAAN	240
QY	241	MVSMANNHMTNSGVSM	300
DB	241	MVSMANNHMTNSGVSM	300
QY	301	LGCGVAANLGRAASVGS	360
DB	301	LGCGVAANLGRAASVGS	360
QY	361	GOMGAGGAGGSLGVL	391
DB	361	GOMGAGGAGGSLGVL	391

Db 361 GQMGARAGGGLSGVLRVPPRPVMPHSPAAG 391

RESULT 2

H70741

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: H70741

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Hollroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID: 98295987; PMID: 9634230

A: Accession: H70741

A: Status: Preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-396 <COL>

A: Cross-references: GB:275555; GB:AL123456; NID:G3261608; PIDN:CAA99966.1; PID:e250360;

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: PPE

Query Match 85.0%; Score 1656.5; DB 2; Length 396;

Best Local Similarity 85.1%; Pred. No. 6.7e-91;

Matches 338; Conservative 18; Mismatches 34; Indels 7; Gaps 3;

QY 1 MVDGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGWSIG 60

Db 1 MVDGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGWSIG 60

QY 61 SSAGLWAAASPYVAVMSVTAGQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120

Db 61 SSAGLWAAASPYVAVMSVTAGQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 120

QY 121 LIATNLGONTPTAVNEAEYGEWMAQDAAMFGYAAATATATATATLTPFEAPMTSAGG 180

Db 121 LIATNLGONTPTAVNEAEYGEWMAQDAAMFGYAAATATATATATLTPFEAPMTSAGG 180

QY 181 LLEQAAVEEASDTAAANQLMNNVPOALQLOAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240

Db 181 LLEQAAVEEASDTAAANQLMNNVPOALQLOAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240

QY 241 MVSNNHMTNSGVSMNTLSMLKGFAPAAAQAVOTAAQNGVRAMSS---LGSSL 296

Db 241 MVSNNHMTNSGVSMNTLSMLKGFAPAAAQAVOTAAQNGVRAMSS---LGSSL 296

QY 297 GSSGLGGVAAANLGRAASVGSLSVPQAAWAAANQAVTPAARALPLTSTSAERPGQMLG 356

Db 297 GSSGLGGVAAANLGRAASVGSLSVPQAAWAAANQAVTPAARALPLTSTSAERPGQMLG 356

QY 357 GLPVQMGARAG--GGLSGVLRVPPRPVMPHSPAAG 391

Db 357 GLPVQMGARAG--GGLSGVLRVPPRPVMPHSPAAG 391

QY 360 GLPLQGLTNSGGGVSNNALRMPRAYVMPRPAAG 396

Db 360 GLPLQGLTNSGGGVSNNALRMPRAYVMPRPAAG 396

RESULT 3

C70568

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: C70568

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Hollroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID: 98295987; PMID: 9634230

A: Accession: C70568

A: Status: Preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-393 <COL>

A: Cross-references: GB:AL020221; GB:AL123456; NID:G3250699; PIDN:CAA17711.1; PID:e125460C

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: PPE

Query Match 40.0%; Score 779.5; DB 2; Length 393;

Best Local Similarity 43.7%; Pred. No. 4.3e-39;

Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGWSIG 61

Db 2 VDFGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGWSIG 61

QY 61 SSAGLWAAASPYVAVMSVTAGQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 121

Db 61 SSAGLWAAASPYVAVMSVTAGQAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMI 121

QY 121 LIATNLGONTPTAVNEAEYGEWMAQDAAMFGYAAATATATATATLTPFEAPMTSAGG 180

Db 121 LIATNLGONTPTAVNEAEYGEWMAQDAAMFGYAAATATATATATLTPFEAPMTSAGG 180

QY 181 LLEQAAVEEASDTAAANQLMNNVPOALQLOAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240

Db 181 LLEQAAVEEASDTAAANQLMNNVPOALQLOAQPTQGTTPSSKLGGLWKTVPSPHSPISN 240

QY 241 MVSNNHMTNSGVSMNTLSMLKGFAPAAAQAVOTAAQNGVRAMSS---LGSSL 296

Db 241 MVSNNHMTNSGVSMNTLSMLKGFAPAAAQAVOTAAQNGVRAMSS---LGSSL 296

QY 297 GSSGLGGVAAANLGRAASVGSLSVPQAAWAAANQAVTPAARALPLTSTSAERPGQMLG 356

Db 297 GSSGLGGVAAANLGRAASVGSLSVPQAAWAAANQAVTPAARALPLTSTSAERPGQMLG 356

QY 357 GLPVQMGARAG--GGLSGVLRVPPRPVMPHSPAAG 391

Db 357 GLPVQMGARAG--GGLSGVLRVPPRPVMPHSPAAG 391

QY 360 GLPLQGLTNSGGGVSNNALRMPRAYVMPRPAAG 396

Db 360 GLPLQGLTNSGGGVSNNALRMPRAYVMPRPAAG 396

QY 122 IATNLLGONTPAIAVNEAEYGEWMAQDAADAFGYAAATAATATATLTPFFPEAPEMTSAGGL 181
DB :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
121 ISTNVFGONTSAIAAEEAOYGEWKAQDSAAAYAVAGSSASASA-VTFPSTPPQIANPTAQ 179
DY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
QY 182 LEQAAAVEREASDTAAA--NOLMNNVPALQOOLAQTGCTTTPSSKLGGLWK----- 229
DY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
DB 180 GTQAAAAVATAAGTAQSLTEMITGCLPNALQSLSPLQQSS-NGFLSWLWOLLFGTNPFT 238
DY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
QY 230 -----TVSPHRSPISNMYVMANNHMSTMNTSGVSMINTLTSSMLKGFPAPAAAAQAQTAA 282
DY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
DB 239 SISALLTDLPQYASFYNTEGLPYFS:GMGNFIQSAKTL-GLIGSAAPAABA-----AA 292
DY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
QY 283 QNGVRAMSSILGSSGLGGGAANLGRAASVGSLSVPOAWAA-ANQAVTPAARALPLT 341
DB :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
DB 293 GDAAKGLPLGGLMG-----GGPVAAGLGNAAASVGKLSVPPVWSGPLPGSVTPGAAPLFVS 348
DY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
QY 342 SLTSAEREGQMGLGPLVCGOMGARAGGGLSGVLRRVPRPVMPHPSPAAG 391
DB :||||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
DB 349 TVSNAEPAAFGSLGGLPL----AGAGGAGAGP-RYGFRPTVMARPFAAG 393
DY :||||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:

RESULT 5
A70932
probable ppe protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C/Accession: A70932
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Rajandream, M.A.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Nature 393, 537-544, 1998
A:Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: A70932
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-409 <COL>
A/Cross-references: GB:AL022021; GB:AL123456; NID:G3250699; PIDN:CAA17729.1; PID:e125461
A/Experimental source: strain H37RV
C/Genetics:
A/Gene: PPE

Query Match 38.5%; Score 751; DB 2; Length 409;
Best Local Similarity 44.7%; Pred. No. 2.2e-37;
Matches 177; Conservative 53; Mismatches 138; Indels 28; Gaps 10;

QY 2 VDFGALPPEINSABRYACPGSASLVAAAQMKDVSADLFSAAAFQSWGLTVGSMIGS 61
DB :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
1 MDFGALPPEINSGRMYPGPGGPLLAAAAMDALAEDYSAAASYGSTIEGLIVAPMWGP 60
DY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
QY 62 SAGLMWAAAAGPYVAAWMSVTAGCAELTAAQVEVAAAAYETAYGLTVPPPVIENRAELMIL 121
DB :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
61 SSITMAAAAVFYVAWISVTACQAQGAQGAQAKIAAGVYETAFAATVPPVIEANRALMLS 120
DY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
QY 122 IATNLLGONTPAIAVNEAEYGEWMAQDAADAFGYAAATAATATATLTPFEAPEMTSAGGL 181
DB :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
121 VATNI FGONT PAIATEAHYAEMWAQDAADAAAMYAGSATA-SQLAPSEPPQTNPST 179
DY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
QY 182 LEQAAAVEREASDTAAA-----NOLMNNVPALQOOLAQTGCTTTPSSKLGGLWKTSVPH 234
DB :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
180 AAQSAVAAQAGAASDDITAGLSQLISLLFSTLQSLA--TTATATSASAG--NDTV--- 232
DY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
QY 235 RSPISNMVSMANNHMSTMNTSGV-----MTNLTSSMLKGFPAPAAAAQAQTAAQNGVRAM 289
DB :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
233 LQSITTILANTGPYSIIIGLGAIPGGWWLTFCQILGLAQNAPFVAAALLGPKAAAAGALSPL 292
DY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
QY 290 SSL-GSSIFGS-SGLGGGVAAAMLGRAASVGSLSVPCAWAAAANOAVTPAARALPLTSLTSA- 346
DB :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
293 APLRGYIGDTIPLGGGTGGIARIAYVGSLSVPCGWAEAAPVWRVASVLPFGTCAPAL 352
DY :|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:
QY 347 AERFGQMGLGPLVCGOMGARAGGGL-----SGVLRV 377

Db 353 AAEAPGALFCEMALSSLAGRALAGTAVRSGAARV 388
 RESULT 6
 H70931
 Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: H70931
 R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987; PMID:9634230
 A:Accession: H70931
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-403 <COL>
 A:Cross-references: GB:AL020201; GB:AL123456; NID:G3250699; PIDN:CAA17728.1; PID:el25461
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: PPE

	Query Match	Score	737.5;	DB 2;	Length	403;
	Best Local Similarity	42.6%	Pred No.	1.4e-36;		
	Matches 181;	Conservative	60;	Mismatches 123;	Indels	61;
	Gaps					12;
Qy	2 VDFCALPPEINSARMYAGFGSASLVAAQMDSVASDLFSAASAFQSVVWGLTVGSWIGS	61				
Dy	5 LDFATLPPPEINSARMYSGAGSAPMLAASAAMHGLSAELRASALSYSVLSTLTGEWHGP	64				
Qy	62 SAGIWMVAASPYVAMWSVTGAQELTAQRVARAAAYETAYGLTPPPPVIAENRAELMIL	121				
Dy	65 ASASNTAAAAPYVAMWSVTVARAQQAERAAAAAYEAFAATVPDPPIEANRAQLMAL	124				
Qy	122 IATNLGONTPAIVNEAEYGEWWAQQDAAMFGYAAATAATATATLPFEAPEMTSAGGL	181				
Dy	125 IATNVLGQNAIPAIAATEAQYAEMWSQDAMMYGYAGASAAAT-QLTPTFTEPVQTNASGL	183				
Qy	182 LEQAAAVEEASDTAAA-----NQLMNNVPOALQLOAQTOGT-----TPSS	222				
Dy	184 AAQSAATAHATGASAGAQQTTLSQIIAAPISVLOGLSSTAAFTAFSGPGLLGIIVSGSS	243				
Qy	223 KLGGWKTVSPHRSPISNMVSNANNHMTNSGVSM-TNTLS---SMLKGFAPAATAQAV	278				
Dy	244 WLDKLWALLDEN-----SNFWNTIASGSLFLPSNTIAPFLGLLGGVAAADAAGDV	293				
Qy	279 QTAQNGVRAMSSGLSSL-----GSSGLGGVVAANLGRAASYGSLSVPOAAAANQAVTP	333				
Dy	294 LGKATSG----GLGGALVAPLGSAGGLGTVAAGLIGNAATVGTLSVPPSWTAAAPLASP	348				
Qy	334 AARAL---PLTSLTAAERGPGQMGLGPLVGQMGARAGGGLSGLVRVPP-----RPYVMFH	386				
Dy	349 LGSALGGTTPMWAPPAPVAVAG-----MPGMFFGTMGQGCGF-----RAVPOYGRFRPFVAR	398				
Qy	387 SPAAG 391					
Dy	399 PPAAG 403					

RESULT 7
 B70931
 Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: B70931
 R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998

Query Match	37.5%;	Score 731;	DB 2;	Length 421;
Best Local Similarity	41.1%;	Pred. NO. 3.5e-36;		
Matches 174;	Conservative	55;	Mismatches 160;	Indels 34;
				Gaps 8;

```

Query Match      36.2%; Score 705; DB 2; Length 408;
Best Local Similarity 41.8%; Pred. No. 1.1e-34;
Matches 182; Conservative 47; Mismatches 134; Indels 72; Gaps 14;

      2 VDFGALPEINSRWYACPGSGASIAVAQAQWDSVASDLFSAASAFQSVWGLTVGSMIGS 61
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      1 MDFGVLPEINSRWYACPGSGPWWAAAWDSUAELGLAAGGYRLAIELTQAYWAGP 60

      62 SAGLWVAAGAAPYVAWMSVTAGQAEILTAQVRVAAAAAYETAYGLTVPPVPVAENRAELMIL 121
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      61 AAASMWAAVTPYVAWLSATAGQAQACGACQAAAAAAAYELAFAMTVPPVVAERALLVAL 120

      122 IATNLLGONTPAIIVNAEAYCEWMAQDAAMAFGYAAAAATATATATLLFPEAPEMTSAGGL 181
      : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      121 VATNFFGONTPAIATAEQAEMWMAQDAAMAAWYAGSAAIAT-ELTPPTAAFPVTTSPAAL 179

      182 LEQAAA-----VEASQSTAAANOLMNN-----VP--CALQO--LAOPTQGTTPSSKLL--- 224

```

```
Db 180 AGQAAATVSTVPPLATTAATVAPQLQLQSLSTSLIPWYSAQQWLAENLLGLTPDNRMIV 239
QY 225 -----GGLWKTVPSPHSRISPMVSMANNHMTNSGVNTVTLSSMLKGFAPAAA 275
Db 240 RLIGISYFDEGL-----LOFEASIAQQAIPGTPGGAG--DSGSSVLDSWGFTIFA 287
QY 276 QAVCTAAQNGVRAMSSL--GSSISLGS-----SGLGGVAAANLGSAAASVGSIS 319
Db 288 -----GPRASPSPVAGGAVGGVQTQPQYMYWALDRESIGGSVSAALGKGSAGSLS 338
QY 320 VPQWAAANQAVTPAARALP--LTSLSAAERPGQMLGGLPVGQMGARAGGSLGVUR 376
Db 339 VPDMAARARWANPAARWRLFGDDVTALRGTAENA---LLRGPFMASAGSTGGGF--VHK 393
QY 377 VPRPVMPHSPAAG 391
Db 394 YGFLAVMQRPFPAG 408
```

RESULT 10

```
B70625
probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70625
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: B70625
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-391 <COL>
A: Cross-references: GB:292539; GB:AL123456; NID: g3261714; PID: e304546;
A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: PPE
```

```
Query Match 36.0%; Score 702; DB 2; Length 391;
Best Local Similarity 42.3%; Pred. No. 1.6e-34;
Matches 166; Conservative 62; Mismatches 150; Indels 14; Gaps 7;
QY 2 VDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFOSVVMGLTVGSWIGS 61
Db 1 MDFGALPPEINSARMYAGAGGPMWAAAGAAWNGLAELGTAAASYEVITRLTTESWMGP 60
QY 62 SAGLMTAAASPYVANKSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMTL 121
Db 61 ASAMVAAAQPYLAWTYTAEEAAHAGSQAMASAAAAYEAYATVPEVVAANRALLAAL 120
QY 122 IATNLGQNTPALAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGGL 181
Db 121 VAINVLGINTPALMATEALYAEWKAQDALAYGYAAAG--AAGMLQLSPSPQTINPGGL 179
QY 182 LEQAAVVEASDTAAANQ-----LMNVPAQLQALQOQTGTTTSSKLGWLKWTVPSPHS 236
Db 180 AAQSAAVGSAATAAVNQVSVADLISSLNVAVSGLASPVTSVLDSTGLSGIITADIALLA 239
QY 237 P--ISNNVMANNHMS--MTNSGVSMNTLSSMLKGFAPAAAQAVCTAAQNGVRAMSSLIG 293
Db 240 TPFVANIINSVNTAAYWYNAIPTALFLANLNSGAPVIAEGATEAAEG---AASAAA 296
QY 294 SSLGSGGLGGVAAANLGRAASVGSLSVPQWAAANQAVTPAARALPLTSLTSAERPGQ 353
Db 297 AGLADSVTPAGLGSALGATLVGRLSVPAASWSTAAPATTAGATALEGSGWTVAAEEA-GP 355
QY 354 MLGGLPVQMGARAGGSLGVLRVPRPVYMP 385
Db 356 VTGMMP-GMAAANKGTGAYAGPRYGRKPTVMP 386
```

RESULT 11

```
B70560
probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70560
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: B70560
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-413 <COL>
A: Cross-references: GB:295436; GB:AL123456; NID: g3261770; PID: CAB08826.1; PID: e316565;
A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: PPE
```

```
Query Match 36.0%; Score 702; DB 2; Length 413;
Best Local Similarity 39.1%; Pred. No. 1.7e-34;
Matches 168; Conservative 61; Mismatches 145; Indels 56; Gaps 7;
QY 1 MVDGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFOSVVMGLTVGSWIG 60
Db 1 MLDFAQLPPEVNSALMYAGPGSGPMLAAWAAWALAELOTTASTYDALITGLADGPWQ 60
QY 61 SSAGLMTAAASPYVANKSVTAGQAEELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMI 120
Db 61 SSSAASVMAAATPQVWLRSTAGQAEQAGSQAVAAASAYEAAFFATVPPEIAANRALLMA 120
QY 121 LIATNLGQNTPALAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGG 180
Db 121 LLATNFGQNTAAATAEYAEWKAQDAAMAYGAGASAAAT--QLSPFPNPAQTINPAG 179
QY 181 LLEQAAVVEASDTAAANQLMNVPAQLQALQOQTGTTTSSKLGWLKWTVPSPHSRISN 240
Db 180 LASQAASVQAVSGAANAQAALTDIPKAL-----FGLSGIFTNEPPPLTLGK 226
QY 241 MVMANNHMTNSGVSMNTLSSMLKGFAPAA---AAQAVQT----- 280
Db 227 ALGLTGHTWSSDGSGLIVGGVLFVQGVTSAGELDSVAMDTFGKVVSPARLMVTFKD 286
QY 281 -----AAQNGVRAMSSLSGSSLSGLGGVAAANLGRAASVGSLSVPQA 323
Db 287 YFGLAHDLPKWASEGAKAAGEAKALPAAVPAIFSAGL--SCVAGAVGQAASVGLKVPV 345
QY 324 WAAANQAVTPAARALPLTSLTSAERPGQMLGGLPVQMGARAGGSLGVLR--RVPRP 381
Db 346 WTATTAPASPAVLAAASNLGAAAAAAGSTHAFGGMPL--MGSGAGAFNNPAPRYGPK 403
QY 382 YVPHSPAAG 391
Db 404 TVIAQPPAGG 413
```

RESULT 12

```
C70931
probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: C70931
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Sgares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
```

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: C70931

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-463 <COL>

A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAAL17723.1; PID:e125462

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match 35.3%; Score 688.5; DB 2; Length 463;

Best Local Similarity 42.8%; Pred. No. 1.3e-33;

Matches 174; Conservative 48; Mismatches 136; Indels 49; Gaps 11;

QY 2 VDFGALPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVWGLTVGWSIGS 61

DB 1 MDFGVLPPEINSGRMYAGPGSGPMLAAAWDGLATELQSTAAADYGSVISVLT-GVWSGQ 59

QY 62 SAGLMVAASPYVAMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMIL 121

DB 60 SSGTMAAAAPYVAMSAALAREAAQAASAAAAYEAAFAATVPFPVVAANRAELAVL 119

QY 122 IATNLGQNTPAIAVNEAEYGEWAQDAAMAFGAAATATATATATLLPFEAPEMTSAGGL 181

DB 120 AATNIFGQNTGALAAEAARYAEMQAQDAAMYGAGSSVAT-QVTFFAAPPTTNAAGL 178

QY 182 LEQAAAVEASDSTAANQLMNNVPAQLQAQTOGTTPPSKLGGLWKTYS--PHRSP1- 238

DB 179 ATQGVAAQVAGSAGN-ARSLVSEVLEFLA--TAGTNYNKTVASLMAVTVGPVYASSVY 235

QY 239 -----SNMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRA 288

DB 236 NSMLGLGPAESKMLPANDTVISTIFGVQPKFFNPTVPNPDLIPK----- 283

QY 289 MSSLGSSLG-----SSGLGG--GVAANLGRAASVGLSVLPQAWAANQVTPAARALPL 340

DB 284 -SALGAGLGLRSATSSGLGSTAPALASAGASQAGSVGMSVPPSWAATPAIRTVAAVFS 342

QY 341 TSLTS--AAERGPQML-----GGLPVQMGARAGGGLSVLRV 377

DB 343 TGLQAVPAAATSEGLLSQMALASVAGGALGAAARATGGLGGGRV 389

RESULT 13

B70932

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: B70932

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: B70932

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-468 <COL>

A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAAL17730.1; PID:e125462

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match 35.0%; Score 682.5; DB 2; Length 468;

Best Local Similarity 40.9%; Pred. No. 2.9e-33;

Matches 164; Conservative 59; Mismatches 125; Indels 53; Gaps 9;

QY 2 VDFGALPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVWGLTVGWSIGS 61

DB 1 MDFGLQPEITSGEWMYLPAGPAGPMLAAAWDGLAELQSWAASVASEVMSWIGP 60

QY 62 SAGLMVAASPYVAMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMIL 121

DB 61 SSAGMAAAAPYVVMWSGTSAQAKAAQAARAAVAYETAFAAVVPPIAARNSQLISL 120

QY 122 IATNLGQNTPAIAVNEAEYGEWAQDAAMAFGAAATATATATATLLPFEAPEMTSAGGL 181

DB 121 VATNIFGQNTAAIAATEAEYGEWAQDTMAEFGYASSATA-SLTPFTAPPQTTPNSGL 179

QY 182 LEQAAAVEAS-----DTAAANQLMNNVPAQLQAQTOGTTPSSKLGGL- 227

DB 180 ACQAAATGQATALASGTNAVITALLSSAAQPPDIIFLLQGLA--TLSTQYTLQMGQII 237

QY 228 -----WKTYSPIHRSPISNNMNNHMTNSGVSMNTLSSMLKGFAPAAAQ 276

DB 238 NAIFGPTGATTYQNVFVTAANVTFSTWANDAMSAPNLGTMTEKFV-----WQPPPAPE 291

QY 277 AVQTAQNGVRAMSSLSGLG-----SSGLGGVAAANLGRAASVGLSVLPQAWAANQAV 331

DB 292 IPK-----SSLGAGLGLRSAGLSAGLAHAASAGLQANLVGDLSPVPSWASATPAV 341

QY 332 TPAARALPLTSLTA-AERGPQMLGGLPVQMGARAGGGL 371

DB 342 RLVAANTLPATSLAAAPATQIPANLLGQWALGSM---TGCAL 379

RESULT 14

A70646

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: A70646

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: A70646

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-380 <COL>

A:Cross-references: GB:Z83867; GB:AL123456; NID:g3261695; PIDN:CAB06278.1; PID:e291015;

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: PPE

Query Match 34.6%; Score 675; DB 2; Length 380;

Best Local Similarity 43.1%; Pred. No. 6.2e-33;

Matches 172; Conservative 56; Mismatches 143; Indels 28; Gaps 12;

QY 2 VDFGALPEINSARMYAGPGSASLVAAQMWDSVASDLFSAASAFQSVWGLTVGWSIGS 61

DB 1 MDFALLPEVNSARMYTGFGAGSLIAAAGGWDLSAAELATTAETAYGSLVSLAALHWGP 60

QY 62 SAGLMVAASPYVAMSVTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMIL 121

DB 61 RAESMAVTAAPYIGWLY-TAEKTOQTALQARAALAFQAVAMTLPPPVAAARIQLAL 120

QY 122 IATNLGQNTPAIAVNEAEYGEWAQDAAMAFGAAATATATATATLLPFEAPEMTSAGGL 181

DB 121 IATNIFGQNTAAIAATEAEYGEWAQDAAMAFGAAATATATATATLLPFEAPEMTSAGGL 179

QY 182 LEQAAAVEASDSTAANQLMNNVPAQLQAQTOGTTPSSKLGGLWKTYSPIHRSP1 238

DB 180 TAQAAAVSQATDPLSL--LIETVTCALQALIPSPFIPEDFTFLDAIPAGYATVGTQDVE 237

QY 239 SNMYSM--ANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGL 296

DB 238 SFVAGTIGAESNLGILNVNDENPAEPTVPGDFGIGELVSA----TSPGGVSA-SGAG--- 289

QY 297 GSSGLGGVAAANLGRAASVGLSVLPQAWAANQAVTPAARALPLTSL--TSAASERG-PGQ 353

Search completed: June 30, 2004, 16:53:58
Job time : 11.2189 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:42:21 ; Search time 6.3659 Seconds
(without alignments)
3199.204 Million cell updates/sec

Title: US-09-597-796c-8

Perfect score: 1949

Sequence: 1 MYDFGALPPEINSARMYAGP.....SGVLRVPRPYWPHSPAAG 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1656.5	85.0	396	1 YD61_MYCTU	Q11031 mycobacteri
2	705	36.2	408	1 YS92_MYCTU	Q10813 mycobacteri
3	698.5	35.3	463	1 Y102_MYCTU	Q53951 mycobacteri
4	444	22.8	447	1 Y402_MYCTU	P42611 mycobacteri
5	426.5	21.9	443	1 Y878_MYCTU	Q10540 mycobacteri
6	418	21.4	678	1 YF48_MYCTU	Q10778 mycobacteri
7	364.5	18.7	408	1 SRA_MYCLE	Q07297 mycobacteri
8	324.5	16.6	463	1 Y095_MYCTU	Q10892 mycobacteri
9	324	16.6	434	1 YU18_MYCTU	P31500 mycobacteri
10	321.5	16.5	435	1 YU21_MYCTU	O53268 mycobacteri
11	233.5	12.0	178	1 Y129_MYCTU	Q06246 mycobacteri
12	217.5	11.2	176	1 Y125_MYCTU	Q50703 mycobacteri
13	153.5	7.9	860	1 ELS_MOUSE	P54320 mus musculu
14	151.5	7.8	232	1 Y126_MYCTU	Q50702 mycobacteri
15	150	7.7	864	1 ELS_RAT	Q09372 rattus norv
16	143.5	7.4	730	1 ELS_HUMAN	P15502 homo sapien
17	139.5	7.2	881	1 FR13_YEAST	P47033 saccharomyc
18	135	6.9	907	1 A180_HUMAN	O60641 homo sapien
19	133.5	6.8	825	1 ICP0_HSV2H	P28284 herpes simp
20	133.5	6.8	2432	1 Y43R_IRV6	P18305 chilo iride
21	132	6.8	1120	1 STFR_ECOLI	P76072 escherichia
22	132	6.8	2090	1 N214_HUMAN	P35658 homo sapien
23	131	6.7	836	1 VG26_BPML5	Q05233 mycobacteri
24	131	6.7	1150	1 APWU_FIG	P22021 sus scrofa
25	129	6.6	1783	1 RAA3_CHLRE	Q9frec4 chlamydomon
26	128	6.6	779	1 SRP_DROME	P52172 drosophila
27	127.5	6.5	790	1 ANP_NOTCO	P24856 notothenia
28	126	6.5	354	1 VAUG_SCHPO	Q10169 schizosacch
29	125.5	6.4	1211	1 BUN2_DROME	Q24523 drosophila
30	125	6.4	577	1 CST2_HUMAN	P33240 homo sapien
31	124	6.4	2090	1 HFC1_MESAU	P51611 mesocricetu
32	123.5	6.3	677	1 Y136_MYCTU	Q50597 mycobacteri
33	123	6.3	432	1 YF10_MYCTU	P71789 mycobacteri

ALIGNMENTS

RESULT 1

ID	YD61_MYCTU	STANDARD;	PRT;	396 AA.
AC	Q11031;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Hypothetical PPE-family protein Rv1361c/MT1406.			
GN	Rv1361C OR MT1406 OR MTCY02B10.25C.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
OX	[1]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=H37RV;			
RC	MEDLINE=98295987; PubMed=9634230;			
RX	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Reltwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Ossegeer K., Skelton S., Squares S., Squares R.,			
RA	Rutter S., Taylor K., Whitehead S., Barrell B.G.;			
RA	Sulston J.B., Taylor K., Whitehead S., Barrell B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RX	MEDLINE=22206494; PubMed=12218036;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,			
RA	Delcher A., Utterback T., Weidman J.J., Khouri H., Gill J., Mikula A.,			
RA	Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;			
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			
RL	J. Bacteriol. 184:5479-5490(2002).			
CC	!- SIMILARITY: belongs to the mycobacterial PPE family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
CC	EMBL; 275555; CAA99966.1; -			
DR	EMBL; AB007013; AAK45669.1; -			
DR	PIR; H70741; H70741.			
DR	TIGR; WTL406; -			
DR	TubercuList; Rv1361C; -			
DR	InterPro; IPR000030; Microbac_PPE.			

Q04893 saccharomyc
P45805 alcaligenes
P35828 caulobacter
Q9ul36 homo sapien
P52591 rattus norv
P10105 drosophila
Q95140 rattus norv
P23314 xanthomonas
P13709 drosophila
P58938 xanthomonas
Q9z798 chlamydia p
Q13492 homo sapien

DR Pfam; PF00823; PPE; 1.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 158 159 TA -> AT (IN REF. 2).
 SQ SEQUENCE 396 AA; 40015 MW; 6AFAD07B5F668D0 CRC64;

Query Match 85.0%; Score 1656.5; DB 1; Length 396;
 Best Local Similarity 85.1%; Pred. No. 1.3e-89;
 Matches 338; Conservative 18; Mismatches 34; Indels 7; Gaps 3;

QY 1 MYDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
 DB 1 MYDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
 QY 61 SSAGLMVAASPYVAMSVTGAQELTAQVRAAAAYETAYGLTPPPVIAENRAELMI 120
 DB 61 SSAGLMVAASPYVAMSVTGAQELTAQVRAAAAYETAYGLTPPPVIAENRAELMI 120
 QY 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLPPFEAPMTSAGG 180
 DB 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLPPFEAPMTSAGG 180
 QY 181 LLEQAAVEEADTAAANQLMNNVPAQLOQAQPTQGTTPSKLGLLWKTVPSPHRSPI 240
 DB 181 LLEQAAVEEADTAAANQLMNNVPAQLOQAQPTQGTTPSKLGLLWKTVPSPHRSPI 240
 QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSS----LGSSL 296
 DB 241 IVSMNNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNGVRAMSS----LGSSL 296
 QY 297 GSSGLGGVAAANLGRAASVGLSVFPQAAWAAANQAVTPAARALPLTSLTSAERPGQMLG 356
 DB 300 GSSGLGGVAAANLGRAASVGLSVFPQAAWAAANQAVTPAARALPLTSLTSAERPGHMLG 359
 QY 357 GLPVQCMQARAG--GGLSGVLRVPRPYVMPHSPAAG 391
 DB 360 GLPLQLGTLNSGGGFGVSNALRMPPRAYVMPVRPAAG 396

RESULT 2
 ID YS92 MYCTU STANDARD; PRT; 408 AA.
 AC Q10813;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein RV2892c/MT2959/MB2916c.
 GN RV2892C OR MT2959 OR MTCY274.23C OR MB2916C.
 OS Mycobacterium tuberculosis, and
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773, 1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL J. Bacteriol. 184:5479-5490(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 CC -/- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; Z74024; CAA98377.1; -;
 CC EMBL; AE007119; AAK47285.1; -;
 CC EMBL; BX248344; CAD96603.1; -;
 CC PIR; G70925; G70925.
 CC TIGR; W2959; -;
 CC Tuberculist; RV2892c; -;
 CC InterPro; IPR000030; Microbac_PPE.
 CC Pfam; PF00823; PPE; 1.
 CC KW Hypothetical protein; Transmembrane; Complete proteome.
 CC FT TRANSMEM 56 76 POTENTIAL.
 CC SQ SEQUENCE 408 AA; 41469 MW; 3E3D1F20D7827199 CRC64;

Query Match 36.2%; Score 705; DB 1; Length 408;
 Best Local Similarity 41.8%; Pred. No. 2.7e-34;
 Matches 182; Conservative 47; Mismatches 134; Indels 72; Gaps 14;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIGS 61
 DB 1 MDGVLPEINSGRMYAGPGSGPMMAAAMDSLAELGLAAGGVRLAISLTGAYWAGP 60
 QY 62 SAGLMVAASPYVAMSVTGAQELTAQVRAAAAYETAYGLTPPPVIAENRAELMI 121
 DB 61 AAASVVAATPVVAVLSATAGCAQAGQAQAAAAAAYELAFAMTVPVPPVYVANNRALLVAL 120
 QY 122 IATNLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLPPFEAPMTSAGGL 181
 DB 121 VATNFFGQNTPAIAATEAQYAEWMAQDAAMVYAGSAIAT-ELTPTTAAPVTTSPAAL 179
 QY 182 LEQAAA-----VEEASDTAAANQLMNN-----VP--QALQO--AQPTQGTTPPSKL--- 224
 DB 180 AQQAATVSVSTVPPIATTAAPVQLQLSLTSLIPWYSAQQWLAENLLGLTPDNRMIV 239
 QY 235 -----GGLWKTVPSPHRSPISNMYSMANNHMTNSGVSMNTLSSMLKGFAPAAAA 275
 DB 240 RLLGYSYFDEGL-----LQFEASLAQAQIPGTPGGAG--DSGSSVLDSWGPPIFA 287
 QY 276 QAVQTAQNGVRAMSSL--GSSLGS-----SGLGGVAAANLGRAASVGLSLS 319
 DB 288 -----GPRASPVAGGAGVGQVTPQPYWYWDRESIGGSVAALKGSSAGSLS 338
 QY 320 VPQWAAANQAVTPAARALP---LTSLSAERPGQMLGCLPVQCMQARAGGGLSGVLR 376
 DB 339 VFPDWAARWANFAAWRLPGDDVTALRGTAENA---LLRGFPNASAGQSTGGGF--VHK 393
 QY 377 VPPRPYVMPHSPAAG 391
 DB 394 YGFRLLAVQVRPPFAG 408

```

RESULT 3
Y102 MYCTU
ID Y102 MYCTU STANDARD; PRT; 463 AA.
AC O53951;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV1802/MT1851/MB1830.
GN RV1802 OR MT1851 OR MT049.24 OR MB1830.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Barrell B.G.;
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umavay L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -1- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC ENBL; AL020201; CAAL1773.1; -
CC ENBL; AE007044; AAK46123.1; -
CC ENBL; BX248340; CAD94533.1; -
CC PIR; C70931; C70931.
CC TIGR; MT1851; -
CC TubercuList; RV1802; -
CC InterPro; IPR000030; Microbac_PPE.
CC Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 401 401 S -> L (IN REF. 2).

```

```

SQ SEQUENCE 463 AA; 46021 MW; EE64828BF09FA551 CRC64;
Query Match 35.3%; Score 688.5; DB 1; Length 463;
Best Local Similarity 42.8%; Pred. No. 2.8e-33;
Matches 174; Conservative 48; Mismatches 136; Indels 49; Gaps 11;
QY 2 VDFGALPPEINARMYAGPGSASIVAAAQWMDSVASDLFSAASAFQSVVWGLTVGWSIGS 61
DB 1 MDFGVLPPEINSGRMVYAGPGSGLMAAAAADWGLATELQSTADYGVISVLT-GWSGQ 59
QY 62 SAGLWAAAAPYVAWMSVTAQQAELTAAQVRVRAAAAYETAYGLTVPPPPVIAENRAELMIL 121
DB 60 SSGTMAAAAAPYVAWMSATLAALFAEAQAASAAAAAYEALFAATVPPVVAANRAELAVL 119
QY 122 IATNLLQNTPAIIVNEAEGEMWAQDAAMFGVAAATATATATALLPFEAPENTSAGGL 181
DB 120 AATNIFQNTGAIAAEARVAEMWAQDAAMYGAGSSVAT-QVTFFAAPPTTNAAGL 178
QY 182 LEQAAVEEAASDTAAANQLMNVPOALQQLAQPTQGTTPSSKLGGLWKTYS--PHRSP1- 238
DB 179 ATQGVAAQVAGASAGN-ARSLVSEVLEFLA--TAGTNYNKTVASLNAVTVGVYASSVY 235
QY 239 -----SNMYSMANNNHSMNTNSGVSMNTLSSMLKGFAPAAAQAQVQTAANGVRA 288
DB 236 NSMLGLGFAESKMWLPANDTVISTIFGMVQPKFFNPVTFNPDLIPK----- 283
QY 289 MSSLGSSIG-----SSGLGG--GVAANLGRAASVGLSVPAQWAAANQAVTPAARALEL 340
DB 284 -SALGAGLGLRSALSSGLGSTAPPAISAGASQAGSVGMSVPPSWAAATPAIRTVAVFSS 342
QY 341 TSLTS--AAERGPQQL-----GGLPVQMGARAGGGLSGVLRV 377
DB 343 TGLQAVFAAAISEGSLLSQALASVAGGALGGAARATGGLGGRV 389
RESULT 4
Y442 MYCTU
ID Y442 MYCTU STANDARD; PRT; 487 AA.
AC P42611; O53727;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV0442C/MT0458.
GN RV0442C OR MT0458 OR MT037.06C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Erdmann;
RX MEDLINE=87137260; PubMed=3029018;
RA Shinnick T.M.;
RT "The 65-kilodalton antigen of Mycobacterium tuberculosis.";
RL J. Bacteriol. 169:1080-1088 (1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Barrell B.G.;
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;

```

```

RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M15467; AA88235.1; ALT INIT.
DR EMBL; AL021932; CAAL7399.1; -
DR EMBL; AE006948; AAK44681.1; -
DR PIR; C70830; C70830.
DR TIGR; MT0458; -.
DR TubercuList; RV0442c; -.
DR InterPro; IPR000303; Microbac_PPE.
DR Pfam; PF01469; Pentapeptide_2; 5.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 40 40 E -> K (IN REF. 2).
FT CONFLICT 96 96 I -> T (IN REF. 1).
FT CONFLICT 211 211 G -> GNNIG (IN REF. 1).
SQ SEQUENCE 487 AA; 47247 MW; 97234D5B316C8C7F CRC64;
Query Match 22.8%; Score 444; DB 1; Length 487;
Best Local Similarity 32.1%; Pred. No. 4.8e-19;
Matches 135; Conservative 48; Mismatches 180; Indels 58; Gaps 12;
QY 4 FGALPPEINSAVMYAGPGSLVAQAQWDSVADLFSAAQFVVMGTVGWSWIGSSA 63
Dd 6 PAWLPPPEINSAVMYAGPGSLVAQAQWDSVADLFSAAQFVVMGTVGWSWIGSSA 65
QY 64 GLMVAAPSPYVAWMSVTAGQELTAQVRVAAYETAYGLTVPPPIAENRAELMLIA 123
Dd 66 AAMVAATQYLAWLUSTAAQAEQAQAQAMATATAFEAALAAATVQPAVVAANRGLMQLIA 125
QY 124 TNLGQNTPAIVNAEYGEWQAQDAAMFGYAAATATATATLPPPEAPEMTSAG--- 179
Dd 126 TNWFGQAPALMDVEAAVEQWALDVAAMAGYHFDASAAVAQLAPMQQV--LNLGIDIG 183
QY 180 --GLLEQAAVEASDTAAANGLMNVLPALQLOLAQTOGTTPSSKLG-----GLW 228
Dd 184 KNGQINLFGNTGSGNIGNNNIGNNNIGNTGNTGNTGNTGNTGNTGNTGNTGNTGNTG 243
QY 229 KTVS-----PHRSPISNWSMANNHSMNTSGVSMNTLSSMLKGPAPAAQAQVQ 279
Dd 244 NTGSGNIGFGITGDHQMGFGFNSGSGN--TGFNSGNTGNTGNTGNTGNTGNTGNTG 287
QY 280 TAAQNGVAMSSLSGSSGLGGVGAANIGRAASVGSLSVPQAWAANAQVTPAAPALP 339
Dd 288 GSGNIGIGSGNSLGSIGTSGT---INAGLG---SAGSLNT-SFWNAGMNAALGAAGS 340
QY 340 LTSLSAARPGQM-----LGLLPVQMGARAG--GGLSSVLR--VPPPEYVMPHSPAA 390
Dd 341 EALVSSAGATGCMSTAAALSSGILASLGTGLQHLANVLNSGLTNPVPAAPASAPV 400
QY 391 G 391
Dd 401 G 401

```

RESULT 5

```

Y878 MYCTU STANDARD; PRT; 443 AA.
ID Q10540;
AC 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV0878c/MT0901.
GN RV0878C OR MT0901 OR MTCY31.06C.
OS Mycobacterium tuberculosis
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Broech R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J.F., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z73101; CA97385.1; -; ALT_INIT.
DR EMBL; AE006977; AAK45143.1; -
DR PIR; C70780; C70780.
DR TIGR; MT0901; -.
DR TubercuList; RV0878c; -.
DR InterPro; IPR000303; Microbac_PPE.
DR InterPro; IPR002989; Mycobac_Pentapep.
DR Pfam; PF01469; Pentapeptide_2; 4.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Transmembrane; Repeat; Complete proteome.
FT TRANSMEM 15 35 POTENTIAL.
FT TRANSMEM 38 58 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 181 201 POTENTIAL.
FT DOMAIN 64 73 POLY-ALA.
FT DOMAIN 81 115 ALA-RICH.
FT DOMAIN 231 270 4 X 10 AA APPROXIMATE REPEATS.
SQ SEQUENCE 443 AA; 43592 MW; C58BEC607F0675E2 CRC64;
Query Match 21.9%; Score 426.5; DB 1; Length 443;
Best Local Similarity 31.2%; Pred. No. 4.5e-18;
Matches 125; Conservative 56; Mismatches 132; Indels 87; Gaps 14;

```


RX MEDLINE=95020554; PubMed=7934845;
 RA Rinke de Wit T.F., Clark-Curtiss J.E., Abebe F., Kolkon A.H.J.,
 RA Jonson A.A.M., Thole J.B.R.;
 RT "A Mycobacterium leprae-specific gene encoding an immunologically
 RT recognized 45 kDa protein.";
 RL Mol. Microbiol. 10:829-838(1993).
 RN [3].
 RP SEQUENCE FROM N.A.
 RC STRAIN=IN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duchaey S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Maclean J., Moulton S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus.";
 RL Nature 408:1007-1011(2001).
 CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U00015; AAC43220.1; -;
 DR EMBL; X68431; CAA48480.1; -;
 DR EMBL; Z21952; CAA79950.1; -;
 DR EMBL; Z97179; CAB09338.1; -;
 DR EMBL; AL583948; CAC29919.1; -;
 DR FIC; C86960; C86960.
 DR FIC; S33522; S33522.
 DR FIC; S39872; S39872.
 DR Leproma; MLO411; -;
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1
 DR Antigen; Repeat; Complete proteome.
 FT DOMAIN 192 196 POLY-SER.
 FT DOMAIN 209 235 2 X 6 AA REPEATS OF S-V-A-Q-S-E.
 FT REPEAT 209 214 1.
 FT REPEAT 230 235 2.
 FT CONFLICT 132 132 T -> S (IN REF. 2).
 FT CONFLICT 189 189 S -> L (IN REF. 2).
 FT CONFLICT 191 191 H -> D (IN REF. 2).
 FT CONFLICT 292 292 P -> L (IN REF. 2).
 SQ SEQUENCE 408 AA; 42466 MW; 500C2BEB0D6E9A9D8 CRC64;
 Query Match 18.7%; Score 364.5; DB 1; Length 408;
 Best Local Similarity 26.3%; Pred. No. 1.7e-14;
 Matches 115; Conservative 72; Mismatches 163; Indels 87; Gaps 11;
 Qy 1 MVDFGALPPEINARMYAGPGSALVAAAQMVDSVADLFSAAAFQSVVGLTVGSGWIG 60
 Db 1 MFDPMVYSPEVNAFLMSRPGSTPLWGAABAWISLABQLMEAAQEVSDTIWVAVPASFAG 60
 Qy 61 SSAGLMAAASPVYAAKSVTGAQELTAAQVRVAARAYETAYGLTVPPIVIAENRAELMI 120
 Db 61 ETSMLASRSTVFVWLDGNAENGLTARVHVAFAFEARAGMVFLLTVLGNIIHTWA 120
 Qy 121 LIATNLGQNTPAIVANAEYGEYMAQDAAMFCYAAATATATATLIPPEAPEMITSAGG 180
 Db 121 LKAINWFGQVSTTVALEADYDLMMWQNSTAMTYTRDTVLRETGKMEPEPAPQLVSR-Y 179
 Qy 181 LLEQAAVAEASDTAAANQLMNNVPQALQQLAQ----- 213
 Db 180 CMRRDVSNGVSHSSSSSDSYESIDNLYDSVAQSEEHGSDMSQSYNTCGSVAQSELCD 239

Qy 214 ----PTQ-----GTPSSKLGGLKWTVPSPHRSPISNMVSMANNHSMNTNSGVSMNTNT 261
 Db 240 PGTPTSQSSQNDLSATSLTQQLGGL-----DSIISSASASLLTTNS--ISSST 286
 Qy 262 LSLMLKGFAPAAAAAQAQAVTAAGNGVRAMSSGLSSGLSGSLGGVAAAN-----LGRAASVG 316
 Db 287 ASSIM-----PIVASQVTEILGRSQV-AVEKMTQSISSSTAVSDVVAASKVAVGAVSGV 341
 Qy 317 SLISVPOAANAANQAVTPAARALP--LTSLSAERPGQMLGGLPVGQMGARAGGLSGV 374
 Db 342 ALRVPENWATASQPVWATAHSPAGCSAITTA-----VSGPLEGV 381
 Qy 375 LVVPRPVVWPHSPAAG 391
 Db 382 TQ--PAEVLITASVAGG 396
 RESULT 8
 Y096 MYCTU STANDARD; PRT; 463 AA.
 AC Q1082;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein Rv0096/MT0105.
 GN Rv0096 OR MT0105 OR MTCV251.15.
 OS Mycobacterium tuberculosis
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI TaxID:1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares R., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.P., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL J. Bacteriol. 184:5479-5490(2002).
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (potential).
 CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z74410; CAA98932.1; -;
 DR EMBL; AE006922; AAK44327.1; -;
 DR FIC; H70750; H70750.
 DR TIGR; MT0105; -;
 DR TuberculList; Rv0096; -

DR	InterPro: IPR000030; Microbac_PPE.	
DR	Pfam: PF00823; PPE; 1.	
KW	Hypothetical protein; Transmembrane; Complete proteome.	
FT	TRANSMEM 3 23	POTENTIAL.
FT	TRANSMEM 88 108	POTENTIAL.
FT	TRANSMEM 112 132	POTENTIAL.
FT	TRANSMEM 216 236	POTENTIAL.
FT	TRANSMEM 245 265	POTENTIAL.
FT	TRANSMEM 275 296	POTENTIAL.
FT	TRANSMEM 323 343	POTENTIAL.
FT	TRANSMEM 419 439	POTENTIAL.
SQ	SEQUENCE 463 AA; 46894 MW; 429D9D66A033D0DD8 CRC64;	
Query Match 16.6%; Score 324.5; DB 1; Length 463;		
Best Local Similarity 27.8%; Pred. No. 4e-12;		
Matches 112; Conservative 54; Mismatches 144; Indels 93; Gaps 12;		
QY	6 ALPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWSSAGL 65	
DB	2 AIPPEVHSGLLSAGCGFGLLVAAQWQELSQQYALACAEQLGQLGGEVQASWQGTAAQ 61	
QY	66 MVAASPYVAMSVTAGQAEITAAQVRVAAAYETAYGLTVPPVIAENRAELMILIATN 125	
DB	62 YVAHGPYLANLEPQTAINSVTRAQHVAAAYCSALAAAMPITPAELAAHNAHGVLIATN 121	
QY	126 LLQONTPAIAVNAEYGEWMAQDAAMFGVAAATATATATLLPFEAPDWTAGGLLEQA 185	
DB	122 FFGINTVPIALNEADYVRWQLAQTMAAYQAVADAATVAVPSTQAPPPIRAPGG---- 176	
QY	186 AAVEEASDT-----AAANOLMNVVQALQOLAQPTQ-----GTPSSKL----- 224	
DB	177 -----DAADTDLVDSIGQIIRI-----LDPIANFYKTFLEFFQFGFPAVTVLVALVAL 229	
QY	225 ---GGLW-----KTVSPHRSPISNMVMNMHMTNSGVSMTNTLSMLK-----GF 269	
DB	230 QLYDFLWYPPYVYGVGLLLLPFFTP-----TLTALTALSAHLINLPPAGL 275	
QY	270 APAAAQAVQTAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVQPAWAAANQ 329	
DB	276 LPFAAA-----LGPQDQGANLAVVATPATAAVP-----GGSP 308	
QY	330 AVTPAPARALPLTSLTAAERGGP---QMLGGLPVG-QMGARAG 368	
DB	309 PTNPAPAAASNSVGSASAPGISVAVPLAPGVSSGPKAG 351	
RESULT 9		
YU18	MYCTU	
ID	YU18_MYCTU STANDARD; PRT; 434 AA.	
AC	P31500; O53265;	
DT	01-JUN-1993 (Rel. 26, Created)	
DT	16-OCT-2001 (Rel. 40, Last sequence update)	
DT	10-OCT-2003 (Rel. 42, Last annotation update)	
DE	Hypothetical PPE-family protein Rv3018c/MT3098/MT3101.	
GN	Rv3018c OR MT3098/MT3101 OR MTV012.32C.	
OS	Mycobacterium tuberculosis.	
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.	
OX	NCBI_TaxID=1773;	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=H37RV;	
RC	MEDLINE=98295987; PubMed=9634230;	
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,	
RA	Gordon S.V., Eigmler K., Gas S., Barry C.E. III, Tekala P.,	
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,	
RA	Davies R., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,	
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,	
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,	
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,	
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;	
RT	"Deciphering the biology of Mycobacterium tuberculosis from the	
RT	complete genome sequence."	
RL	Nature 393:537-544 (1998).	
RL	[2]	
RN	SEQUENCE FROM N.A.	
RP	STRAIN=CDC 1551 / Oshkosh;	
RC	MEDLINE=22208494; PubMed=12218036;	
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,	
RA	Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,	
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,	
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,	
RA	Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;	
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and	
RT	laboratory strains";	
RL	J. Bacteriol. 184:5479-5490 (2002).	
RL	[3]	
RN	SEQUENCE OF 160-374 FROM N.A.	
RP	STRAIN=Isolate 50410;	
RC	Patki A.H., Dale J.W.;	
EL	Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.	
CC	-!- SIMILARITY: Belongs to the mycobacterial ppe family.	
CC	-!- CAUTION: In strain Oshkosh the gene for this protein is	
CC	interrupted in position 307 by an 18610 element.	
CC	-!- CAUTION: Was originally (Ref.3) thought to be a dihydrofolate	
CC	reductase.	
CC	-!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts	
CC	in positions 294; 337 and 355.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; AL021287; CA416103.1; -	
DR	EMBL; AE007129; AAK47427.1; ALT_SEQ.	
DR	EMBL; AE007129; AAK47430.1; ALT_SEQ.	
DR	EMBL; X59271; CAA41961.1; ALT_FRAME.	
DR	PIR; E70857; E70857.	
DR	TIGR; MT3098; -	
DR	TIGR; MT3101; -	
DR	TubercuList; RV3018c; -	
DR	InterPro; IPR000030; Microbac_PPE.	
DR	Pfam; PF00823; PPE; 1.	
KW	Hypothetical protein; Complete proteome.	
SQ	SEQUENCE 434 AA; 43029 MW; 41D673C4BD389DD6 CRC64;	
Query Match 16.6%; Score 324; DB 1; Length 434;		
Best Local Similarity 28.0%; Pred. No. 4e-12;		
Matches 115; Conservative 56; Mismatches 178; Indels 62; Gaps 12;		
QY	6 ALPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWSSAGL 65	
DB	8 ASPPEVHSGALLSAGPGSLQAAAGNSALSAEYAAVQELSUVVAAVAGVWQGPSAEL 67	
QY	66 MVAASPYVAMSVTAGQAEITAAQVRVAAAYETAYGLTVPPVIAENRAELMILIATN 125	
DB	68 FVAAYVPVAVLVQASADSAAAAGHEAAAAGYVCAEAEMPTLPBLAAHNLTHAVLVATN 127	
QY	126 LLQONTPAIAVNAEYGEWMAQDAAMFGVAAATATATATLLPFEAPDWTAGGLLEQA 185	
DB	128 FFGINTVPIALNEADYVRWVQAAATVNGAYAVVGAALVATPHTGPAPVIVKPG----- 181	
QY	186 AAVEEASDTAAAN-----QLMNVVQALQOLAQPTQGTTPSSKLGSLKWTVPSPH 234	
DB	182 --ANEASNAVAATITPPWHEIVQLEETFAAYDQYLSALLSELPA--VAWVWQPLFVD 237	
QY	235 ---RSPISNMVSMANNHMTNSGVSMTNTLSMLKGA----- 270	
DB	238 ILGFNIIGTIIITLASNQAQLTEFAINASYVAVGLLYATAGVIDIVVWVIGNLFGVWPLL 297	
QY	271 -----PAAAQAVQTAAQNGVRAMSSLGSSGLGGVAAANLGRAASV--GSLSVPOA 323	


```
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;  
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and  
RL J. Bacteriol. 184:5479-5490(2002)."  
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; Z95389; CAB08678.1; -;  
DR EMBL; AB007158; AAK47873.1; -;  
DR PIR; C70975; C70975.  
DR TIGR; MT3533; -;  
DR TubercuList; RV3429; -;  
DR InterPro; IPR000030; Microbac_PPE.  
DR Pfam; PF00823; PPE; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 178 AA; 19811 MW; 8BE1FC025ABFBEA6 CRC64;  
  
Query Match 12.0%; Score 233.5; DB 1; Length 178;  
Best Local Similarity 35.9%; Pred. No. 2.8e-07;  
Matches 65; Conservative 25; Mismatches 76; Indels 15; Gaps 3;  
  
QY 7 LPPEINARMYAGPGASLVAAQAQWDSVAFQSVVWGLTVGSMI 59  
DB 5 IPAEYISNIIIEGPGADSLFASQGLRELAYSVETTAESLEDELD-ENWKGSSDDL 63  
QY 60 GSSAGLWVAASPYVAMSVTAGQALTAQVRAAAAYETAYGLTVPPVIAENRAELM 119  
DB 57 GSSSDLWADAAGLYDLWTKHSQILETAVIDFLAVVYEETRHKVVPATIANNEEVH 116  
QY 120 ILIATNLGONTPAIIVNAEYEGEMWAQDAAMFGYAAATATATATALLPFEAPMTSAG 179  
DB 117 RLIIASNVAGVNTPAIAGLDAQYQYRAQNTAVNVNDYQSTARFILAYLPRWQEPPIYGGG 176  
  
QY 180 G 180  
DB 177 G 177  
  
RESULT 12  
ID_VY25_MYCTU STANDARD; PRT; 176 AA.  
AC Q50703;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hypothetical PPE-family protein RV3425.  
GN RV3425 OR MTCV78.04C.  
OS Mycobacterium tuberculosis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.  
OX NCBI_TaxID=1773;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=H37Rv;  
RX MEDLINE=98295987; PubMed=9634230;  
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,  
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Teika F.,  
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,  
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,  
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,  
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,  
RA Ruter S., Seeger K., Skelton S., Squares R., Squares R.,  
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;  
RT "Deciphering the biology of Mycobacterium tuberculosis from the  
RL complete genome sequence."  
RL Nature 393:537-544(1998).
```

```
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; Z77165; CAB01031.1; -;  
DR PIR; F70738; F70738.  
DR TubercuList; RV3425; -;  
DR InterPro; IPR000030; Microbac_PPE.  
DR Pfam; PF00823; PPE; 1.  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 176 AA; 19855 MW; B8CEFE29463B87B0 CRC64;  
  
Query Match 11.2%; Score 217.5; DB 1; Length 176;  
Best Local Similarity 33.5%; Pred. No. 2.3e-06;  
Matches 58; Conservative 31; Mismatches 83; Indels 1; Gaps 1;  
  
QY 7 LPPEINARMYAGPGASLVAAQAQWDSVAFQSVVWGLTVGSMI 59  
DB 5 IPAEYISNIIIEGPGADSLFASQGLRELAYSVETTAESLEDELD-ENWKGSSDDL 63  
QY 67 VAAASPYVAMSVTAGQALTAQVRAAAAYETAYGLTVPPVIAENRAELMILATNL 126  
DB 64 ADAVERYLQWLSKHSQKHAAWVINGLANAYNDRKRWVPEPIAANREERRLLASNV 123  
QY 127 LGONTPTAVNAEYEGEMWAQDAAMFGYAAATATATATALLPFEAPMTSAG 179  
DB 124 AGVNTPAIADLQAQYQYRAQNTAVNVNDYQSTARFILAYLPRWQEPPIYGGG 176  
  
RESULT 13  
ID_ELS_MOUSE STANDARD; PRT; 860 AA.  
AC F54320;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 01-OCT-1996 (Rel. 34, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Elastin precursor (Tropoelastin).  
GN ELN.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALE/C; TISSUE=Lung;  
RX MEDLINE=95133069; PubMed=7429060;  
RA Wyder K.S., Sechler J.L., Boyd C.D., Passmore H.C.;  
RT "Use of an intron polymorphism to localize the tropoelastin gene to  
RT mouse chromosome 5 in a region of linkage conservation with human  
RT chromosome 7."  
RL Genomics 23:125-131(1994).  
CC -!- FUNCTION: Major structural protein of tissues such as aorta and  
CC nuchal ligament, which must expand rapidly and recover completely.  
CC -!- SUBUNIT: The polymeric elastin chains are cross-linked together  
CC into an extensible 3D network.  
CC -!- SUBCELLULAR LOCATION: Extracellular matrix of elastic fibers.  
CC -!- PTM: The crosslinks are made of deaminated Lys.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
CC EMBL; U08210; AAA80155.1; -;
```


Qy	5	GALP----	PEINSARMYAGPGSASILVAAAAQMWDSVADLFSAAAFOSVTVGLTVGSW--	58
Db	399	GGTGGVGGPGIGGPGVLGGPGAVSPAAAKAAKAAK--	YGARGGVGPIPTYGVGAGGPG	456
Qy	59	--IGSSAGLMLVAAAAGPYVAMNSVTGAQELTAAQVRAAAAAAYETAYGLTVPVPIAENRA	116	
Db	457	YGVGACAGT---	GGASORAAA--AAAAKAKYKAG--	486

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:01 ; Search time 28.479 Seconds
(without alignments)
4331.879 Million cell updates/sec

Title: US-09-597-796c-8

Perfect score: 1949

Sequence: 1 WVDPCALPEINSARMYAGP.....SGVLRVPRPYMPHPSPAAG 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_25.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_viruse.*
- 16: sp_bacteriaph.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1949	100.0	391	16	O05298 mycobacteri
2	1920.5	98.5	330	16	Q7U0E9 mycobacteri
3	1779.5	91.3	336	16	Q7U071 mycobacteri
4	1648	84.6	393	16	O8VI23 mycobacteri
5	1583	81.2	393	16	O06341 mycobacteri
6	1583	81.2	393	16	Q7W4F5 mycobacteri
7	779.5	40.0	393	16	Q7T2J3 mycobacteri
8	779.5	40.0	411	16	O53939 mycobacteri
9	754	38.7	410	2	Q99Q11 mycobacteri
10	751	38.5	409	16	O53957 mycobacteri
11	749	38.4	409	16	Q7T2H7 mycobacteri
12	740.5	38.0	399	16	Q7T2H8 mycobacteri
13	737.5	37.8	403	16	O53956 mycobacteri
14	736	37.8	423	16	O53950 mycobacteri
15	736	37.8	423	16	Q7T2I4 mycobacteri
16	731	37.5	421	16	Q9Z5K0 mycobacteri

17	702	36.0	391	16	P96362 mycobacteri
18	702	36.0	391	16	Q7U0T5 mycobacteri
19	702	36.0	413	16	O06386 mycobacteri
20	683.5	35.1	694	16	Q8V0W0 mycobacteri
21	682.5	35.0	468	16	O53958 mycobacteri
22	675	34.6	380	16	P95190 mycobacteri
23	675	34.6	380	16	Q7TX66 mycobacteri
24	670.5	34.4	385	16	Q7T2R7 mycobacteri
25	669.5	34.4	394	16	Q7TXX5 mycobacteri
26	668.5	34.3	462	16	O3310 mycobacteri
27	667.5	34.2	385	16	O3304 mycobacteri
28	666.5	34.2	385	16	Q8VJZ0 mycobacteri
29	641	32.9	364	16	Q7T2J5 mycobacteri
30	637	32.7	350	16	Q7T2J2 mycobacteri
31	635	32.6	363	16	O53940 mycobacteri
32	633.5	32.5	365	16	O86373 mycobacteri
33	629.5	32.3	405	16	Q8VJW5 mycobacteri
34	626.5	32.1	381	16	Q7TX67 mycobacteri
35	618.5	31.7	397	2	Q9AGF0 mycobacteri
36	615.5	31.6	382	16	Q7TXX3 mycobacteri
37	615.5	31.6	402	16	O3312 mycobacteri
38	610	31.3	443	16	O8VKL9 mycobacteri
39	604	31.0	443	16	Q7U242 mycobacteri
40	603	30.9	426	16	Q7U114 mycobacteri
41	603	30.9	426	16	O05907 mycobacteri
42	597.5	30.7	391	16	O05798 mycobacteri
43	594.5	30.5	391	16	Q7TX76 mycobacteri
44	589	30.2	406	16	P71869 mycobacteri
45	589	30.2	406	16	Q7TW99 mycobacteri

ALIGNMENTS

RESULT 1

O05298 ID O05298 PRELIMINARY; PRT; 391 AA.

AC O05298;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein (PPE family protein).
GN RV1196 OR MTC1364.08 OR M1234.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37Rv.
RX MEDLINE=98299987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CDC 1551 / Ohkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.


```

QY 181 LLEQAAAVEASDTAAANLNNVPOALQOLQAQPTGTTSSKLGGLWKTSPHRSPI 240
DB 181 LLEQAAAVEASDTAAANLNNVPOALQOLQAQPTGTTSSKLGGLWKTSPHRSPI 240
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAONGVRAMSS-----LGS 296
DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAONGVRAMSS-----LGS 296
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAARPGQM 356
DB 300 GSSGLGGVAAANLGRAASVGSLSVPPAWAANAQVTPAARALPLTSLTSAARPG 359
QY 357 GLPVGOMGARAG--GGLSGVLVPPRPYPVPHSPAG 391
DB 360 GLPLGOLTNSGGGFGVSNALRMPPRAYVMPVPAAG 396

RESULT 4
Q8VIZ3
ID Q8VIZ3 PRELIMINARY; PRT; 393 AA.
AC Q8VIZ3;
DT 01-MAR-2002 (T-EMBLrel. 20, Created)
DT 01-MAR-2002 (T-EMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE PPE family protein.
GN MT3582.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB007161; AAK47941.1; -.
DR TIGR; MT3582; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR PROSITE; PS00501; SPASE I.1; 1.
SQ SEQUENCE 393 AA; 39688 MW; 86F0B67798855511 CRC64;

Query Match 84.8%; Score 1648; DB 16; Length 393;
Best Local Similarity 84.8%; Pred. No. 2e-85;
Matches 395; Conservative 20; Mismatches 34; Indels 6; Gaps 3;

QY 1 MVDFGALPPINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSM 60
DB 1 MVDFGALPPINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSM 60
QY 61 SSAGLMVAASAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAASAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLGONTPTAIVNEAEYGEWQAQDAAMFGYAAATATATATLPPPEAPMTSAG 180
DB 121 LIATNLGONTPTAIVNEAEYGEWQAQDAAMFGYAAATATATATLPPPEAPMTSAG 180
QY 181 LLEQAAAVEASDTAAANLNNVPOALQOLQAQPTGTTSSKLGGLWKTSPHRSPI 240
DB 181 LLEQAAAVEASDTAAANLNNVPOALQOLQAQPTGTTSSKLGGLWKTSPHRSPI 240
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAONGVRAMSS-----LGS 296
DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAONGVRAMSS-----LGS 296
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAARPGQM 356
DB 297 GSSGLGGVAAANLGRAASVGSLSVPPAWAANAQVTPAARALPLTSLTSAARPG 359

```

```

DB 241 IVSMNNHVMNTNSGVSMNTLSSMLKGFAP-AAAQAQVTAONGVRAMSS-----LGS 299
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAARPGQM 356
DB 300 GSSGLGGVAAANLGRAASVGSLSVPPAWAANAQVTPAARALPLTSLTSAARPG 359
QY 357 GLPVGOMGARAG--GGLSGVLVPPRPYPVPHSPAG 391
DB 360 GLPLGH-SVNAGSGINNLRVPAVAIPTPAAG 393

RESULT 5
C06341
ID C06341 PRELIMINARY; PRT; 393 AA.
AC C06341;
DT 01-JUL-1997 (T-EMBLrel. 04, Created)
DT 01-JUL-1997 (T-EMBLrel. 04, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Hypothetical protein RV3478.
GN RV3478 OR MTCY113E12.31.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
DR EMBL; Z95390; CAB08702.1; -.
DR PIR; C70568; C70568.
DR Tuberculist; RV3478; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;

Query Match 81.2%; Score 1583; DB 16; Length 393;
Best Local Similarity 81.5%; Pred. No. 9.3e-82;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 1 MVDFGALPPINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSM 60
DB 1 MVDFGALPPINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSM 60
QY 61 SSAGLMVAASAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
DB 61 SSAGLMVAASAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLGONTPTAIVNEAEYGEWQAQDAAMFGYAAATATATLPPPEAPMTSAG 180
DB 121 LIATNLGONTPTAIVNEAEYGEWQAQDAAMFGYAAATATATLPPPEAPMTSAG 180
QY 181 LLEQAAAVEASDTAAANLNNVPOALQOLQAQPTGTTSSKLGGLWKTSPHRSPI 240
DB 181 LLEQAAAVEASDTAAANLNNVPOALQOLQAQPTGTTSSKLGGLWKTSPHRSPI 240
QY 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAONGVRAMSS-----LGS 296
DB 241 VSSIANNHSMNTNSGVSMNTLSSMLKGLAP-AAAQAQVTAONGVRAMSS-----LGS 299
QY 297 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAARPGQM 356

```

```

Db 300 GSSGLGAGVAANLGRAASVSGSLVPPAWAANQAQVTPAARALPLTSLTSAATAPGHMLG 359
QY 357 GLPVGQMGARAGGSLGVLVRPPPPVPMHPSPAAG 391
Db 360 GLPLGH-SVNAGSGINNLRVPARAYAIPTPTPAAG 393

RESULT 6
Q7TWFS PRELIMINARY; PRT; 393 AA.
ID Q7TWFS
AC Q7TWFS
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DE PE family protein.
DE PP260 OR ME3505.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97; PubMed=12788972;
RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248346; CAD9562.1; -.
KW Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF420C95DAE7DD4 CRC64;

Query Match 81.2%; Score 1583; DB 16; Length 393;
Best Local Similarity 81.5%; Pred. No. 9.3e-82;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWIG 60
Db 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWIG 60
QY 61 SSAGLWAAASPYVAMSVTAGQAELETAQVRVAAAYETAYGLTVPPVIAENRAELMI 120
Db 61 SSAGLWAAASPYVAMSVTAGQAELETAQVRVAAAYETAYGLTVPPVIAENRAELMI 120
QY 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 180
Db 121 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 180
QY 121 LTATNLLGQNTPAIEANQAAVSQWKGDAEAMYGAAATATATATALLPFEDALITNPGG 180
Db 121 LTATNLLGQNTPAIEANQAAVSQWKGDAEAMYGAAATATATATALLPFEDALITNPGG 180
QY 181 LLEQAAVEEASDTAAANQLMNNVPOALQLOAQPTQCTTSSKLGKWLKTVSPHRSPIN 240
Db 181 LLEQAAVEEASDTAAANQLMNNVPOALQLOAQPTQCTTSSKLGKWLKTVSPHRSPIN 240
QY 181 LLEQAAVEEASDTAAANQLMNNVPOALQLOAQPTQCTTSSKLGKWLKTVSPHLSPLN 240
Db 181 LLEQAAVEEASDTAAANQLMNNVPOALQLOAQPTQCTTSSKLGKWLKTVSPHLSPLN 240
QY 241 MVSNNHMSMTNGSVNTLTSLMLKGFAPAAAQVQTAAGNGVRAMS----LCSSSL 296
Db 241 VSSANNHMSMTNGSVNTLTSLMLKGLAP-AAAQAVETAENGWVWMSLSGSLGSSL 299
QY 297 GSSGLGGVVAANLGRAASVSGSLVPPAWAANQAQVTPAARALPLTSLTSAAGRGPMGLG 356
Db 297 GSSGLGGVVAANLGRAASVSGSLVPPAWAANQAQVTPAARALPLTSLTSAAGRGPMGLG 356
QY 300 GSSGLGAGVAANLGRAASVSGSLVPPAWAANQAQVTPAARALPLTSLTSAATAPGHMLG 359
Db 300 GSSGLGAGVAANLGRAASVSGSLVPPAWAANQAQVTPAARALPLTSLTSAATAPGHMLG 359
QY 357 GLPVGQMGARAGGSLGVLVRPPPPVPMHPSPAAG 391
Db 360 GLPLGH-SVNAGSGINNLRVPARAYAIPTPTPAAG 393

RESULT 7
Q7TWJ3 PRELIMINARY; PRT; 393 AA.
ID Q7TWJ3
AC Q7TWJ3
DT 01-OCT-2003 (TReMBLrel. 25, Created)

```

```

DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE PPE family protein.
DE PPE26 OR ME1817.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97; PubMed=12788972;
RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248340; CAD94520.1; -.
KW Complete proteome.
SQ SEQUENCE 393 AA; 38572 MW; 41944B6E547A8AB0 CRC64;

Query Match 40.0%; Score 779.5; DB 16; Length 393;
Best Local Similarity 43.7%; Pred. No. 2e-36;
Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;

QY 2 VDFGALPPEINSARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSGWIGS 61
Db 1 MDGFGALPPEVNSVRMYAGPGSAPMVAASAWNGLAAELSSAATGVETITQLSSEGMLGP 60
QY 62 SAGLWAAASPYVAMSVTAGQAELETAQVRVAAAYETAYGLTVPPVIAENRAELMI 121
Db 61 ASAGLWAAASPYVAMSVTAGQAELETAQVRVAAAYETAYGLTVPPVIAENRAELMI 120
QY 122 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGGL 181
Db 121 ISTNVFGQNTSAIAAAEAQYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGGL 179
QY 182 LEQAAVEEASDTAAANQLMNNVPOALQLOAQPTQCTTSSKLGKWLKTVSPHRSPIN 229
Db 180 GTQAAVAATAAGTAGTSTLTEMITGLPNAQLSTLPLQSS-NGPLSLMLWQLFOTPNPPT 238
QY 230 -----TVSPHRSPISNNVSMANNHMSMTNGSVNTLTSLMLKGFAPAAAQVQTA 282
Db 239 SISALLTDLQYAFSEFFNYTEGLPYFSGKGNFQAAKTL-GLIGSAAAPA-AAAA 292
QY 283 QNGVAMSSLSGSSGLGGVVAANLGRAASVSGSLVPPAWAANQAQVTPAARALPLT 341
Db 293 GDAAGLPLGLGMLG----GGPVAAGLGNAAASVGKLSVPPVWMSGPLPGSVTPGAAPLVS 348
QY 342 SITSAAERPGOMLGLGVLPGQMGARAGGSLGVLVRPPVPMHPSPAAG 391
Db 349 TVSAAPAAAGSLGGLPL-----AGAGGAGAGP-RYGRPTVMARPPFAG 393

RESULT 8
OS3939 PRELIMINARY; PRT; 411 AA.
ID OS3939
AC OS3939
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE PPE-family protein.
GN RV1789 OR MT1838 OR MTV049.11.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,

```

```
RA Gorden S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajadream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fletschmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey B.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL020201; CAAL7711.1; ALT_INIT.
DR EMBL; AB007043; AAK46108.1; -.
DR F1R; G70929; G70929.
DR TIGR; MT1838; -.
DR TubercuList; Rv1789; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Complete proteome.
SQ SEQUENCE 411 AA; 40557 MW; 03C90B5E059087DA CRC64;

Query Match 40.0%; Score 779.5; DB 16; Length 411;
Best Local Similarity 43.7%; Pred. No. 2.1e-36;
Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;

QY 2 VDFGALPPPEINSGARMYAGPGSASLVAAQWDSVSDLSFAASAFQSVVWGLTVGSWIGS 61
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 19 MDFGALPPPEINSGARMYAGPGSAPVMAAAGWGLAELSSAATGVTITQLSSSEGWLGP 78
QY 62 SAGLVAAASPVYVWMSVTAQAEELTAQVRAAAAYETAYGLTVPPPPVIAENRAELMI 121
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 79 ASNMAEAVPYVWMSAAQAQAEQATQARAAAAPFAATVPPPLIANNRASLMQL 138
QY 122 IATNLGONTPAIVNAEYEGEMWQAQDAAMFGYAAATATATATLPPFEAPMTSAGL 181
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 139 ISTNVFGQNTSAIAAAEAQYEGEMWQAQDAAMFGYAAATATATATLPPFEAPMTSAGL 197
QY 182 LEOQAAVEEASDTAAA--NQLMNNVPOALQOAOPTQGTTPSSKLGGLWK----- 229
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 198 GTQAAAATAAGTAQSTLTETMITGLPNALQSLTSLPQLSS--NGPLSWLMQILFGTPNFFT 256
QY 230 -----TVSPHRSPISSNMVSMANNHMSMTNSGV---SMTNTLSSMLKGFAPAAAQAVQTAA 282
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 257 SISALLTDIQPYASFFYNTGELPYFSGIGNFNFIQSAKTL-GLIGSAPAAVA-----AA 310
QY 283 QNGVRAVSSIGSSGLSGGVAANLGRAASVGLSVPOAWAA--ANQAVTPAARALPLT 341
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 311 GDAAKGLPGLGGLMG----GGPVAAGLGNAAVSGKLSVPPVMSGGLPFGSVTTCGAAPLVS 366
QY 342 SLTSAERGGOMLGLPVQMGARAGGLSGVLRVPPPYVPHSPHAPAG 391
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 367 TVSAAPAPGSLGGLPL----AGAGGAGAGP-RYGFRPTVWARPPFAG 411

RESULT 9
Q99Q11 PRELIMINARY; PRT; 410 AA.
AC Q99Q11; 2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE Rv1808-like protein.
GN MYC1808 OR OV1808.
```

```
OS Mycobacterium microti.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1806;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MYC 94-2272, and OV254;
RA Sivadon V., Heym B., Mazancourt P., Gaillard J.-L.L.;
RA "PPE Rv1808 orthologue of Mycobacterium microti.";
RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AF335180; AAK20894.1; -.
DR EMBL; AF335179; AAK20893.1; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
SQ SEQUENCE 410 AA; 40044 MW; 3B9DA3174655A5EA CRC64;

Query Match 38.7%; Score 754; DB 2; Length 410;
Best Local Similarity 44.5%; Pred. No. 5.7e-35;
Matches 177; Conservative 51; Mismatches 140; Indels 30; Gaps 10;

QY 1 MVDFGALPPPEINSGARMYAGPGSASLVAAQWDSVSDLSFAASAFQSVVWGLTVGSWIG 60
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 1 MLDGALPPPEINSGARMYAGPGSGFLLAANAADALAAELYSAAASYGSTIEGLTVAPWVG 60
QY 61 SSAGLVAAASPVYVWMSVTAQAEELTAQVRAAAAYETAYGLTVPPPPVIAENRAELMI 120
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 61 PSSITMAAAVAPYVWISVTAGQAEQAGAKIAAGVYETAFATVPPPPVIAENRALLMS 120
QY 121 LIATNLGONTPAIVNAEYEGEMWQAQDAAMFGYAAATATATATLPPFEAPMTSAGG 180
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 121 LVATNIFGQNTPAIVNAEYEGEMWQAQDAAMFGYAAATATATATLPPFEAPMTSAGG 179
QY 181 LLEQAAVEEASDTAAA-----NQLMNNVPOALQOAOPTQGTTPSSKLGGLWKTVP 233
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 180 TAAQSAVVAAQAGAAAGSSDITAQLSLIPSTLQSLA--TTATATSAGS--WDTV-- 233
QY 234 HRSPISSNMVSMANNHMSMTNSGV---SMTNTLSSMLKGFAPAAAQAVQTAAQNGVRAMS 290
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 234 -LQSIITILANLTGPGYSIILGLAIPGGMWLTFGGIL-GLAQNAPGVAAALGPKAAAGALS 291
QY 291 SLGSSLSG-----SGLGGVGAANLGRAASVGLSVPOAWAAANQAVTFAARALFLTSLTS 345
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 292 PLALRGYIADITPLGGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 351
QY 346 A-AERGGOMLGLPVQMGARAGGL-----SGVLRV 377
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
DQ 352 ALAAEAFGALFEGMALSSLAGALAGTAVRSAGAAARV 389

RESULT 10
OS3957 PRELIMINARY; PRT; 409 AA.
AC OS3957;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-JUN-1998 (TReMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE PPE-family protein.
GN Rv1808 OR MT1856.1 OR MTV049.30.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garner T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajadream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
```

RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RL complete genome sequence.";
 RN Nature 393:537-544 (1998).
 RP [2]
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL022021; CAAL7729.1; -
 DR EMBL; AE007044; AAK46129.1; ALT_INIT.
 DR PIR; A70932; A70932.
 DR TIGR; MT1856.1; -
 DR TubercuList; Rv1808; -
 DR InterPro; IPR000030; Microbac_PPB.
 DR Pfam; PF00823; PPE; 1.
 KW Complete proteome.
 SQ SEQUENCE 409 AA; 39917 MW; 1E15202BACF36379 CRC64;

Query Match 38.5%; Score 751; DB 16; Length 409;
 Best Local Similarity 44.7%; Pred. No. 8.4e-35;
 Matches 177; Conservative 53; Mismatches 138; Indels 28; Gaps 10;

QY 2 VDFGALPPEINSARMYAGPGSGASLVAAQWMDVSADLFSASAFQSVVMGLTVGWSIGS 61
 DB 1 MDFGALPPEINSGRMYAGPGSGPLLAARAAWDAALAEALYSAAASYGSTIEGLTVAPWMPG 60
 QY 62 SAGLMVAASPYVAMSVTAGOAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMIL 121
 DB 61 SSITMAAAVAPYVAMISVTAGAEQAQAQAKIAAGVYETAFATVPPVIEANRALLMSL 120
 QY 122 IATNLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLLPFEAPBMTSAGGL 181
 DB 121 VATNIFGONTFAIAATEAHYAEMWQAQDAAMYGAGSSATA-SQLAPFSEPPQTTFNSAT 179
 QY 182 LEQAAVVEASDTHAA-----NQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTYSHPH 234
 DB 180 AAQSAVVAQAAGAAASDDITAQLSQLISLLPSTLQSLA--TTATATSASAG--WDTV--- 232
 QY 235 RSPISNMVSMNNHMTNSGV-----MTNTLSMLKGFAPAAQAQVTAQAQNGVRAM 289
 DB 233 LQSIITILANTGPYSIIIGLGAIPGGWHLTFQILGLAQNAPGVAALLGPKAAGALSPL 292
 QY 290 SSL-GSSLGS-SGLGGVAAANLGRAASVGLSVPOQAAANQAVTPAARALPITSLSA- 346
 DB 293 APLRGYIGDITPLGGGATGAIARIVGSLSVPOGWAEAPVWRAVASVLPCTGAAPAL 352
 QY 347 AERPGQMLGCLPVQMGARAGGL-----SGVLRV 377
 DB 353 AAAPGALFGEALSSLAGRALAGTAVRSGAARV 389

RESULT 11

Q7TZH7
 ID Q7TZH7 PRELIMINARY; PRT; 409 AA.
 AC Q7TZH7;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE PPE family protein.
 GN PPE32 OR MB1837.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eighmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
 DR EMBL; BX248340; CAD94540.1; -
 KW Complete proteome.
 SQ SEQUENCE 409 AA; 39931 MW; D57892628B131A9E CRC64;

Query Match 38.4%; Score 749; DB 16; Length 409;
 Best Local Similarity 44.3%; Pred. No. 1.1e-34;
 Matches 176; Conservative 51; Mismatches 140; Indels 30; Gaps 10;

QY 2 VDFGALPPEINSARMYAGPGSGASLVAAQWMDVSADLFSASAFQSVVMGLTVGWSIGS 61
 DB 1 MDFGALPPEINSGRMYAGPGSGPLLAARAAWDAALAEALYSAAASYGSTIEGLTVAPWMPG 60
 QY 62 SAGLMVAASPYVAMSVTAGOAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMIL 121
 DB 61 SSITMAAAVAPYVAMISVTAGAEQAQAQAKIAAGVYETAFATVPPVIEANRALLMSL 120
 QY 122 IATNLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLLPFEAPBMTSAGGL 181
 DB 121 VATNIFGONTFAIAATEAHYAEMWQAQDAAMYGAGSSATA-SQLAPFSEPPQTTFNSAT 179
 QY 182 LEQAAVVEASDTHAA-----NQLMNNVPQALQQLAQPTQGTTPSSKLGGLWKTYSHPH 234
 DB 180 AAQSAVVAQAAGAAASDDITAQLSQLISLLPSTLQSLA--TTATATSASAG--WDTV--- 232
 QY 235 RSPISNMVSMNNHMTNSGV-----MTNTLSMLKGFAPAAQAQVTAQAQNGVRAM 291
 DB 233 LQSIITILANTGPYSIIIGLGAIPGGWHLTFQILGLAQNAPGVAALLGPKAAGALSPL 291
 QY 292 LGSLSGS-----SGLGGVAAANLGRAASVGLSVPOQAAANQAVTPAARALPITSLSA 346
 DB 292 LAPLGGYIGDITPLGGGATGAIARIVGSLSVPOGWAEAPVWRAVASVLPCTGAAPA 351
 QY 347 AERPGQMLGCLPVQMGARAGGL-----SGVLRV 377
 DB 352 AAAPGALFGEALSSLAGRALAGTAVRSGAARV 389

RESULT 12

Q7TZH8
 ID Q7TZH8 PRELIMINARY; PRT; 399 AA.
 AC Q7TZH8;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE PPE family protein.
 GN PPE31 OR MB1836.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eighmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsempe C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
 DR EMBL; BX248340; CAD94539.1; -
 KW Complete proteome.
 SQ SEQUENCE 399 AA; 38840 MW; 1A0F4377318E74F2 CRC64;

Query Match

38.0%; Score 740.5; DB 16; Length 399;

Best Local Similarity 42.8%; Pred. No. 3.2e-34;					
Matches 182; Conservative 60; Mismatches 122; Indels 61; Gaps 13					
QY	2	VDFGALPPEINSARMYAGPGSASLVAAQAQMWDVSADLFSASAFAOSVVMGLTVGSWTGS	61		
		: :			
Db	1	MDFATLPPEINSARMYSGAGSPMLAAASAHWGLSAELRASALSYSVLSTLTGBEWHGP	60		
QY	62	SAGLMVAASPYVANKSVTAGAELTAAQVRVAAAAYETAYGLTVPPPIAENRAELMTL	121		
		: :			
Db	61	ASASMTAAAPYVAVSWVTAVRAOAGAQAFAAAAAAYEAFAATVPFPVIEATRAQLMAL	120		
QY	122	IATNLLQGNTPAIAVNEAEYEMMAQDAAMAFGYAAATATATATLTPFEAPEMTSAGGL	181		
		: :			
Db	121	IATNVLGONAPAIAATEAQYAEMWSQDAMAMYAGASAAAT-QLTPFTPEPVQTINASSL	179		
QY	182	LQQAARVEASTAAA-----NQLMNNVPOALLOLAOPTGGTT---PSKLG-----	225		
		: :			
Db	180	AACSRAINAHATCASAGAQTTLQSOLIAIPSVLQGLSSSTAATSAGPGSLGLILSGSS	239		
QY	226	---GLWKVTSVPHRSPISNNVSMANNHMSTMNTSGVSM-TNTLS---SMLKGFAFAAAAAQAV	278		
		: :			
Db	240	WLDKLWALLDPN-----SNFWNTIASGGLFLPSNTIAPFLGLGGVAAADAAGDV	289		
QY	279	QTAQNGVNAMSGSSSL-----GSSGLGGGVAAANLGRAASVGSLSVPOMAAANAQAVTP	333		
		: :			
Db	290	LGEATSG-----GLGALVAPIGSGAGGLGGTVTAAGLGNAAATVGTLSVPPSWTAAAPLASP	344		
QY	334	AAEAL---PLTSLTSAABEGPCQMLGGLPVGOMGARAGGGLSGVLRVPP----RPVWMPH	386		
		: :			
Db	345	LSGALGCTPMVAPPAPPAVAAAG-----MPGNPFCTMGCGGFG-----RAVPOYGRFRPNFVAR	394		
QY	387	SPAAG 391			
		: : :			
Db	395	PPAAG 399			
RESULT 13					
O53956	PRELIMINARY; PRT; 403 AA.				
ID	O53956				
AC	O53956;				
DT	01-JUN-1998 (T-EMBLrel. 06, Created)				
DT	01-JUN-1998 (T-EMBLrel. 06, Last sequence update)				
DT	01-JUN-2003 (T-EMBLrel. 24, Last annotation update)				
DE	PEE-family protein.				
GN	RV1807 OR MT1856 OR MTV049.29.				
OS	Mycobacterium tuberculosis.				
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1773;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=H37RV;				
RC	MEDLINE=98295987; PubMed=9634230;				
RA	Cole S.T., Brosch R., Parkhill J.J., Garnier T., Churcher C., Harris D.,				
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,				
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Davies R., Devlin K., Feltwell I., Gentles S., Hamlin N., Holroyd S.,				
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,				
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,				
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,				
RA	Sultston J.E., Taylor K., Whitehead S., Barrall B.G.;				
RT	"deciphering the biology of Mycobacterium tuberculosis from the				
RL	complete genome sequence."				
RL	Nature 393:537-544 (1998).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CDC 1551 / Oshkosh;				
RC	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,				
RA	Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,				
RA	Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,				
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,				
RA	Biswal W.;				
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and				

RT	laboratory strains.";
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL; AL022021; CAA17728.1; -.
DR	EMBL; AE007044; AAK46128.1; -.
DR	PIR; H70931; H70931.
DR	TIGR; MT1856; -.
DR	TubercuList; Rv1807; -.
DR	InterPro; IPR000030; Microbac_PPE.
DR	Pfam; PF00823; PPE; 1.
KW	Complete proteome.
FT	CONFLICT 227 F -> S (IN REF. 2).
FT	CONFLICT 238 V -> L (IN REF. 2).
FT	CONFLICT 238
FT	SEQUENCE 403 AA; 39243 MW; DCEI8880FD15CBEE CRC64;
ST	
Query Match	37.8%; Score 737.5; Length 403;
Best Local Similarity	42.6%; Pred. No. 4.9e-34;
Mismatches	181; Conservative
Gaps	61; Indels 123; Mismatches 123; Gaps 123;
Qy	2 VDFGALPPEINSARMYAGPGSASIVAAAQWDSVASDLFSAASAFQSVMGLTYSMTGS 61
Dd	5 LDFATLPPEINSARMYSGAGSAPMLAASAASHHGLSAELRASALSUSVSLTLTGEEWHGP 64
Qy	62 SAGLMVAASAPFYVAMSVTAQGAEHTAAQVRVAAAAAYETAYGTITVPFPVFAENRAEMLIL 121
Dd	65 ASASNTAAAAAPFYVAMSVTVARAEQAQAEAAAAAYEAFAATVPPVIEANRAQLMAL 124
Qy	122 IATNLLGONTPTAIANTAEYGEWMQAADAAWFGVAAAATATATATALLPPEAPEMTSAGGL 181
Dd	125 IATNVLGQNAPATAATEAQYAENMSQDAMMYGYAGASAAAT-OLTFTEPFVQTNNASGL 183
Qy	182 LEQAAAVEEASDTAAA-----NOLMNVVPAQLLOLAQPTCGT-----TPSS 222
Dd	184 AAQGAALAHHAIGASAGAQQTTLSLIIRIIPVLOGLSSTTAATFASGPSGLLVGSGSS 243
Qy	223 KLGGLKWTISPHRSPISNMVSMANNHMSTMNSGVSM-TNTLS----SMLKGPAAPAAAQAV 278
Dd	244 WLDKWALLDPN-----SNFMNTIASGGFLFPSNTIAPPGLILGGVAAAADAAGDV 293
Qy	279 QTAQNGVRAMSISLGSII-----GSSGLGGVVAANLGRAASVGSLSVPAQAAAAAQAVP 333
Dd	294 LGPEATSG-----GUGGALVAPLGSAGGGLGTIVAAGLGNAAITVGLISVPPSPSWTAAPLASP 348
Qy	334 AARAL---PLTSITSAAERPGQMGLGPLVGOMGARAGGGLSGVLRVPPP-----RPYNMH 386
Dd	349 LGSALGGETPMVAPPVPAVAG-- --MPGMWPTMTMGQGFG-----RAVPOYGFRPNFVAR 398
Qy	387 SPAAG 391
Dd	399 PPAAG 403

RESULT 14	PRELIMINARY;	PRT; 423 AA.
OS3950		
ID	OS3950	
AC	OS3950;	
DT	01-JUN-1998 (TrEMBLrel. 06, Created)	
DT	01-JUN-1998 (TrEMBLrel. 06, Last sequence update)	
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)	
DE	PPS-family protein.	
GN	RV1801 OR MTW049.23 OR MT1850.	
OS	Mycobacterium tuberculosis.	
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;	
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.	
OX	NCBI_TaxID=1773;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=H37RV;	
RX	MEDLINE=98295987; PubMed=9634230;	
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,	
RA	Tekaia F.,	
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III,	
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,	
RA	Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,	
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2004, 16:01:22 ; Search time 8830 Seconds
(without alignments)
11225.993 Million cell updates/sec
Title: US-09-597-796c-11
Perfect score: 2287
Sequence: 1 tctagaataatttgttta.....ggntgtacaaagccgaaa 2287

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0
Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :	GenEmbl.*
1:	gb_ba.*
2:	gb_hgt.*
3:	gb_in.*
4:	gb_om.*
5:	gb_ov.*
6:	gb_pat.*
7:	gb_ph.*
8:	gb_pl.*
9:	gb_pr.*
10:	gb_ro.*
11:	gb_sts.*
12:	gb_sv.*
13:	gb_un.*
14:	gb_vt.*
15:	em_ba.*
16:	em_fun.*
17:	em_hum.*
18:	em_in.*
19:	em_mu.*
20:	em_om.*
21:	em_or.*
22:	em_ov.*
23:	em_pat.*
24:	em_ph.*
25:	em_pl.*
26:	em_ro.*
27:	em_sts.*
28:	em_un.*
29:	em_vi.*
30:	em_htg_hum.*
31:	em_htg_inv.*
32:	em_htg_other.*
33:	em_htg_mus.*
34:	em_htg_pin.*
35:	em_htg_rod.*
36:	em_htg_mam.*
37:	em_htg_vrt.*
38:	em_sy.*
39:	em_htgo_hum.*
40:	em_htgo_mus.*
41:	em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2284	99.9	2287	6	BD251322 Fused pro
2	2284	99.9	2287	6	AR303127 Sequence
3	2284	99.9	2287	6	AR403735 Sequence
4	1770.8	77.4	1737	6	BD251333 Fused pro
5	1770.8	77.4	1737	6	AR403746 Sequence
6	1172.8	51.3	1801	6	BD251331 Fused pro
7	1172.8	51.3	1801	6	AR403744 Sequence
8	1169	51.1	1188	6	BD273807 Sequence
9	1169	51.1	1188	6	AX004989 Sequence
10	1169	51.1	3058	6	AR169205 Sequence
11	1169	51.1	3058	6	AR182495 Sequence
12	1169	51.1	3058	6	AR194978 Sequence
13	1169	51.1	3058	6	AR233150 Sequence
14	1169	51.1	3058	6	AR353355 Sequence
15	1169	51.1	3058	6	AX429698 Sequence
16	1169	51.1	3058	6	AX832683 Sequence
17	1169	51.1	3058	6	BD006378 Compounds
18	1169	51.1	3058	6	BD006498 Compounds
19	1169	51.1	3058	6	BD205870 Compounds
20	1169	51.1	15264	1	AE007000 Mycobacte
21	1169	51.1	75216	6	AX704275 Sequence
22	1169	51.1	349306	15	Bx842575 Mycobacte
23	1168.6	51.1	1176	6	BD273806 Sequence
24	1168.6	51.1	1176	6	AX004987 Sequence
25	1164.2	50.9	1565	6	BD069338 Compounds
26	1149.8	50.3	299450	1	BX248338 Mycobacte
27	1021	44.6	299450	1	BX248338 Mycobacte
28	955	41.8	20276	1	AE007161 Mycobacte
29	947.4	41.4	14029	1	AE007013 Mycobacte
30	942.6	41.2	348264	1	BX842576 Mycobacte
31	941	41.1	3027	6	AR169207 Sequence
32	941	41.1	3027	6	AR182497 Sequence
33	941	41.1	3027	6	AR194880 Sequence
34	941	41.1	3027	6	AR233152 Sequence
35	941	41.1	3027	6	AR353357 Sequence
36	941	41.1	3027	6	AX429702 Sequence
37	941	41.1	3027	6	AX832687 Sequence
38	941	41.1	3027	6	BD06380 Compounds
39	941	41.1	3027	6	BD06500 Compounds
40	941	41.1	3027	6	BD205872 Compounds
41	923.8	40.4	316050	1	BX248346 Mycobacte
42	922.2	40.3	349606	15	Bx842583 Mycobacte
43	920.2	40.2	1179	6	AX750742 Sequence
44	884	38.7	1725	6	AR169206 Sequence
45	884	38.7	1725	6	AR182496 Sequence

ALIGNMENTS

RESULT 1	BD251322	2287 bp	DNA	linear	PAT 17-JUL-2003
BD251322	Fused protein of Mycobacterium tuberculosis antigen and utilization thereof.				
LOCUS	BD251322				
DEFINITION	BD251322.1 GI:33061092				
ACCESSION	JP 2002510494-A/1.				
VERSION	synthetic construct				
KEYWORDS	artificial sequences.				
SOURCE	1 (bases 1 to 2287)				
ORGANISM	Skeiky, Y.A.W., Alderson, M. and Neto, A.C.				
REFERENCE	Fused protein of Mycobacterium tuberculosis antigen and utilization thereof				
AUTHORS					
TITLE					

JOURNAL	Patent: JP 2002510494-A 1 09-APR-2002;
COMMENT	CORIXA CORP
	OS Artificial Sequence
	PN JP 2002510494-A/1
	PD 09-APR-2002
	PF 07-APR-1999 JP 2000542460
	PR 07-APR-1998 US 09/056556,30-DEC-1998 US 09/223040 PI
	YASIR A W SKEIKY, MARK ALDERSON, ANTONIO CAMPOS NETO PC
	C12N15/09, A61K39/04, A61K48/00, A61P31/04, C07K14/35, C07K19/00, PC
	C12P21/02,
	PC C12N15/00
	CC Description of Artificial Sequence: tri-fusion protein Ral2- CC
	TDH9-Ra35
	CC (designated Mtb32a)
	CC n = g, a, c or t
	CC n = g, a, c or t
	CC n = g, a, c or t
	CC Key Location/Qualifiers
	FT modified base (30)
	FT modified_base (33)
	FT CDS (42).. (2231)
	FT modified base (2270).
	FT Location/Qualifiers
FEATURES	1. .2287
source	/organism="synthetic construct"
	/mol_type="genomic DNA"
	/db_xref="taxon:32630"
ORIGIN	
	Query Match 99.9%; Score 2284; DB 6; Length 2287;
	Best Local Similarity 100.0%; Pred. No. 7.6e-273;
	Matches 2287; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 TCTAGAATAATTTTGTACTTAAAGANGANATATACATATGCATCACCATCACCATC 60
DB	1 TCTAGAATAATTTTGTACTTAAAGANGANATATACATATGCATCACCATCACCATC 60
QY	61 ACACGGCCGGCTCCGATAAATCTCCAGCTGTCCAGAGTGGCGAGGATTCGCCATTCCGA 120
DB	61 ACACGGCCGGCTCCGATAAATCTCCAGCTGTCCAGAGTGGCGAGGATTCGCCATTCCGA 120
QY	121 TCGGGCAGGGCAGTGGCGATCGCGGGCCAGATCCGATCGGGTGGGGGTCAACCACCGTTTC 180
DB	121 TCGGGCAGGGCAGTGGCGATCGCGGGCCAGATCCGATCGGGTGGGGGTCAACCACCGTTTC 180
QY	181 ATATCGGGCTACCGCTTCTCGGCTTGGGTGTGTGCAACAAACGCGCAACGGCGCAC 240
DB	181 ATATCGGGCTACCGCTTCTCGGCTTGGGTGTGTGCAACAAACGCGCAACGGCGCAC 240
QY	241 GAGTCCAAACGCTGGTTCGGAGCGCTCCGGCGCAAGTCTCGGCATCTCCACCGGCGAGC 300
DB	241 GAGTCCAAACGCTGGTTCGGAGCGCTCCGGCGCAAGTCTCGGCATCTCCACCGGCGAGC 300
QY	301 TGATCACCAGCTGAGGGCTCCGATCAACTCGGCCACCGGATGGCGGACGGCTTA 360
DB	301 TGATCACCAGCTGAGGGCTCCGATCAACTCGGCCACCGGATGGCGGACGGCTTA 360
QY	361 ACGGGCATCATCCCGGTGAGTCACTCGGTGACCTGGCAAAACCAAGTGGGCGGCACGC 420
DB	361 ACGGGCATCATCCCGGTGAGTCACTCGGTGACCTGGCAAAACCAAGTGGGCGGCACGC 420
QY	421 GTACAGGAACGTGACATTTGGCCAGGAGACCCCGCCCGAATTCAATGTTGTTTCGGG 480
DB	421 GTACAGGAACGTGACATTTGGCCAGGAGACCCCGCCCGAATTCAATGTTGTTTCGGG 480
QY	481 CGTTACACCGGAGATCAATCTCCGCGAGGATGACCGCGCGCGGTTCGGCTCGCTGG 540
DB	481 CGTTACACCGGAGATCAATCTCCGCGAGGATGACCGCGCGCGGTTCGGCTCGCTGG 540
QY	541 TGGCCGGCTCAGATGTGGGAACGGTGGCGAGTGAATCTGTTTCGGCCGGTGGGGT 600
DB	541 TGGCCGGCTCAGATGTGGGAACGGTGGCGAGTGAATCTGTTTCGGCCGGTGGGGT 600
QY	601 TTCAGTCGGTGTCTCGGGTCTGACGTCGGGTCTGTGATAGGTTCTGTCGGCGGTCTGA 660
DB	601 TTCAGTCGGTGTCTCGGGTCTGACGTCGGGTCTGTGATAGGTTCTGTCGGCGGTCTGA 660
QY	661 TGGTGGCGGCGCTCGCGTATGTGGCGTGTGATGAGCGTCAACCGCGGGGAGCGCCGAGC 720
DB	661 TGGTGGCGGCGCTCGCGTATGTGGCGTGTGATGAGCGTCAACCGCGGGGAGCGCCGAGC 720
QY	721 TGACCGCGCGCCAGGTCCGGGTGTGTCGGCGGCTACGACGCGGTATGGGTGACGG 780
DB	721 TGACCGCGCGCCAGGTCCGGGTGTGTCGGCGGCTACGACGCGGTATGGGTGACGG 780
QY	781 TGCCCCCGCGGTGTATCCCGAGAACCGTGTGAATGATGATTCGATACGACCAACC 840
DB	781 TGCCCCCGCGGTGTATCCCGAGAACCGTGTGAATGATGATTCGATACGACCAACC 840
QY	841 TCTTGGGGCAAAACACCCCGCGATCGCGGTCAACAGGCGCAATACGGCGAGATGTGGG 900
DB	841 TCTTGGGGCAAAACACCCCGCGATCGCGGTCAACAGGCGCAATACGGCGAGATGTGGG 900
QY	901 CCCAAGACGCGCGCGATGTTTGGCTACGCGCGGGAACGCGACGCGCGACGCGACGCGT 960
DB	901 CCCAAGACGCGCGCGATGTTTGGCTACGCGCGGGAACGCGACGCGCGACGCGACGCGT 960
QY	961 TGCTGCCGTTTGAGGAGCGCGGAGATGACGAGCGGGTGGGCTCTCGAGCAGCGCG 1020
DB	961 TGCTGCCGTTTGAGGAGCGCGGAGATGACGAGCGGGTGGGCTCTCTGAGCAGCGCG 1020
QY	1021 CCGCGTTCGAGGAGCGCTCCGACACCGCGCGGGAACCAAGTTGATGAACAATGTGCC 1080
DB	1021 CCGCGTTCGAGGAGCGCTCCGACACCGCGCGGGAACCAAGTTGATGAACAATGTGCC 1080
QY	1081 AGGCGCTCAACAGTGGCGCGCGACGCGAGGCAACGCGCTTCTTCAAGTGGGTG 1140
DB	1081 AGGCGCTCAACAGTGGCGCGCGACGCGAGGCAACGCGCTTCTTCAAGTGGGTG 1140
QY	1141 GCCTGTGGAAGACGCTCTCGCGCATCGTCCGCGATCAGCAACATGTTGCGATGCCA 1200
DB	1141 GCCTGTGGAAGACGCTCTCGCGCATCGTCCGCGATCAGCAACATGTTGCGATGCCA 1200
QY	1201 ACAACCATGTCATGACCAACTCGGGTGTGTCGATGACCAACACTTGAAGTCGATGT 1260
DB	1201 ACAACCATGTCATGACCAACTCGGGTGTGTCGATGACCAACACTTGAAGTCGATGT 1260
QY	1261 TGAAGGCTTTTGTCCCG 1320
DB	1261 TGAAGGCTTTTGTCCCG 1320
QY	1321 TCCGGCGATGAGTCTCGTGGGAGCTCGTGGGTCTTTCGGGTCTCGGGGTGGGGTGG 1380
DB	1321 TCCGGCGATGAGTCTCGTGGGAGCTCGTGGGTCTTTCGGGTCTCGGGGTGGGGTGG 1380
QY	1381 CCGCCAACTTGGGTTCGGCGCGCTCGGTGGTTCGTTGTCGTCGCGAGCGCTGGCGCG 1440
DB	1381 CCGCCAACTTGGGTTCGGCGCGCTCGGTGGTTCGTTGTCGTCGCGAGCGCTGGCGCG 1440
QY	1441 CGGCAACACGAGTCACTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
DB	1441 CGGCAACACGAGTCACTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
QY	1501 CCGCGGAAGAGGCG 1560
DB	1501 CCGCGGAAGAGGCG 1560
QY	1561 GGGCGGTGTGGGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
DB	1561 GGGCGGTGTGGGTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
QY	1621 ATTCTCGCGAGCGCGCGATATCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1680
DB	1621 ATTCTCGCGAGCGCGCGATATCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1680
QY	1681 TCCCGCGCTGCCCTCGACCCCGTCGCGATGTCGCCCAAGTGGGCGCCACAGGTGGTCA 1740

[illegible]

Db	2221	CCGCGTCTTAGGATATCCATCATCACTGGCGCGCTCGAGCAGATCCGGNTGTACAAAG	2287	2287 bp	DNA	linear	PAT 18-DEC-2003
Qy	2281	CCCGAAA	2287				
Db	2281	CCCGAAA	2287				
RESULT 3							
LOCUS	AR403735						
DEFINITION	Sequence 1 from patent US 6627198.						
ACCESSION	AF403735						
VERSION	AR403735.1	GI:40151411					
KEYWORDS	Unknown.						
SOURCE	Unknown.						
ORGANISM	Unclassified.						
REFERENCE	1 (bases 1 to 2287)						
AUTHORS	Reed S.G., Skeiky Y.A., Dillon, D.C., Alderson, M. and Campos-Neto, A.						
TITLE	Fusion proteins of Mycobacterium tuberculosis antigens and their						
JOURNAL	Patent: US 6627198-A 1 30-SEP-2003;						
FEATURES	Location/Qualifiers						
source:	1..2287						
	/organism="unknown"						
ORIGIN	/mol_type="genomic DNA"						
Query Match	99.9%; Score 2284; DB 6; Length 2287;						
Best Local Similarity	100.0%; Pred. No. 7.6e-273;						
Matches 2287; Conservative	0; Mismatches 0; Indels 0; Gaps 0;						
Qy	1	TCTAGAAATAATTTTCTTTACTTTAAGAANGANATATACATATGCATCACCATCACCATC	60				
Db	1	TCTAGAAATAATTTTCTTTACTTTAAGAANGANATATACATATGCATCACCATCACCATC	60				
Qy	61	ACACGGCCGGCTCCGGATAACTTCCAGCTGTCCAGGGTGGGAGGATTCGCCATTCCGA	120				
Db	61	ACACGGCCGGCTCCGGATAACTTCCAGCTGTCCAGGGTGGGAGGATTCGCCATTCCGA	120				
Qy	121	TCGGCAGAGCGATGGCGATCGGGGCCAGATCGATCGGGTGGGGGTACCCACCGTTTC	180				
Db	121	TCGGCAGAGCGATGGCGATCGGGGCCAGATCGATCGGGTGGGGGTACCCACCGTTTC	180				
Qy	181	ATATCGGGCTACCGCTTCTCGGTCTGGTGTGTGCACAAACACGGCAACGGCGCAC	240				
Db	181	ATATCGGGCTACCGCTTCTCGGTCTGGTGTGTGCACAAACACGGCAACGGCGCAC	240				
Qy	241	GAGTCCAAACGGTGTGGGAGCGCTCCGGCGCAAGTCTCGGCATCTCAACGGCGAAG	300				
Db	241	GAGTCCAAACGGTGTGGGAGCGCTCCGGCGCAAGTCTCGGCATCTCAACGGCGAAG	300				
Qy	301	TGATCACCGGGTCAGCGGCTCCGATCAACTCGGCCACCGGATGGCGAGCGGCTTTA	360				
Db	301	TGATCACCGGGTCAGCGGCTCCGATCAACTCGGCCACCGGATGGCGAGCGGCTTTA	360				
Qy	361	ACGGGATCATCCGGTGTGACGTATCTCGGTGACCTGGCAACCAAGTCGGCGGACGC	420				
Db	361	ACGGGATCATCCGGTGTGACGTATCTCGGTGACCTGGCAACCAAGTCGGCGGACGC	420				
Qy	421	GTCAGGGACGTGACATTTGGCGAGGACCCCGGCGAAATTCATGTGGATTTGGGG	480				
Db	421	GTCAGGGACGTGACATTTGGCGAGGACCCCGGCGAAATTCATGTGTGGATTTGGGG	480				
Qy	481	CGTTACACCGAGATCAACTCCGCGAGATGTACGCCCGCGCGGTTCCGGCTCGCTGG	540				
Db	481	CGTTACACCGAGATCAACTCCGCGAGATGTACGCCCGCGCGGTTCCGGCTCGCTGG	540				
Qy	541	TGGCCCGGCTCATGTGGGACAGGTGGGAGTACCTGTTTTCGGCCGGTGGCGCT	600				
Db	541	TGGCCCGGCTCATGTGTGGGACAGGTGGGAGTACCTGTTTTCGGCCGGTGGCGCT	600				
Qy	601	TTCACTCGGTGTCTGGGGTCTGACGGTGGGGTCTGTGGATAGTTCGTGGCGGGTCTGA	660				

Db 601 TTCAAGTCGGTCTGGGGCTTACCGTGGGGTCTGGATAGTTCTGCGCGGCTCTGA 660
QY 661 TCGTGGCGCGCCCTCGCGCTATGTGGGTGATGAGCGTCAACCGCGGGGAGCCGAGC 720
Db 661 TGGTGGCGCGCCCTCGCGCTATGTGGGTGATGAGCGTCAACCGCGGGGAGCCGAGC 720
QY 721 TGACCGCGCCCGAGTCTCGGGTCTGGCGGGCTTACGAGCGCGTATGGGTGACGG 780
Db 721 TGACCGCGCCCGAGTCTCGGGTCTGGCGGGCTTACGAGCGCGTATGGGTGACGG 780
QY 781 TGCACCGCGCGGTATCGCGAGACCGTGTCTGATGATCTGATAGCGACCAACC 840
Db 781 TGCACCGCGCGGTATCGCGAGAACCGTGTCTGATGATCTGATAGCGACCAACC 840
QY 841 TCTTGGGGCAAAACCCCGCGGATCGCGGTCAACGAGCGCGAATACGCGAGATGTGG 900
Db 841 TCTTGGGGCAAAACCCCGCGGATCGCGGTCAACGAGCGCGAATACGCGAGATGTGG 900
QY 901 CCCAAGACCGCGCGATTTGGCTACCGCGCGGACCGCGACCGCGACCGCGAGT 960
Db 901 CCCAAGACCGCGCGATTTGGCTACCGCGCGGACCGCGACCGCGACCGCGAGT 960
QY 961 TGTCTCGGTTTCAGGAGGCGCGGAGATGACCGAGCGCGGTCTCTCGAGCAGGCGG 1020
Db 961 TGTCTCGGTTTCAGGAGGCGCGGAGATGACCGAGCGCGGTCTCTCGAGCAGGCGG 1020
QY 1021 CCGCGTTCAGAGAGCGCTCCGACACCGCGCGGCGAACCAGTGTATGACATGTGCC 1080
Db 1021 CCGCGTTCAGAGAGCGCTCCGACACCGCGCGGCGAACCAGTGTATGACATGTGCC 1080
QY 1081 AGCGGTCAACAGCTGGCGCCAGCCACCGAGGACCGAGCGCTTCTTCAAGCTGGGTG 1140
Db 1081 AGCGGTCAACAGCTGGCGCCAGCCACCGAGGACCGAGCGCTTCTTCAAGCTGGGTG 1140
QY 1141 GCCTGTGGAAGACGGTCTCGCGCATCGGTCCCGATCAGCAACATGTGTGATGGCA 1200
Db 1141 GCCTGTGGAAGACGGTCTCGCGCATCGGTCCCGATCAGCAACATGTGTGATGGCA 1200
QY 1201 ACAACCATGTGATGACCACTCGGTGTCTGATGACCAACACCTTGACTCGATGT 1260
Db 1201 ACAACCATGTGATGACCACTCGGTGTCTGATGACCAACACCTTGACTCGATGT 1260
QY 1261 TGAAGGGCTTTGTCCGGGGGGCGCGCGAGCGGTGCAAAACCGCGCGCAAAACGGGG 1320
Db 1261 TGAAGGGCTTTGTCCGGGGGGCGCGCGAGCGGTGCAAAACCGCGCGCAAAACGGGG 1320
QY 1321 TCGGGCGATAGCTCGTGGGAGCTCGCTGGGTTCTTGGGTCTGGCGTGGGGTGG 1380
Db 1321 TCGGGCGATAGCTCGTGGGAGCTCGCTGGGTTCTTGGGTCTGGCGTGGGGTGG 1380
QY 1381 CGCCCAACTTGGTTCGGCGGCTCGGTCTGGTTCGGTTCGGTTCGGTTCGGTTCGG 1440
Db 1381 CGCCCAACTTGGTTCGGCGGCTCGGTCTGGTTCGGTTCGGTTCGGTTCGGTTCGG 1440
QY 1441 CGGCCAACAGGACGTCACCCCGCGCGCGCGCTCGGTTCGGTTCGGTTCGGTTCGG 1500
Db 1441 CGGCCAACAGGACGTCACCCCGCGCGCGCGCTCGGTTCGGTTCGGTTCGGTTCGG 1500
QY 1501 CGCGGAAAGAGGCGCGGAGATCTGGCGGGTCTGGGTTCGGTTCGGTTCGGTTCGG 1560
Db 1501 CGCGGAAAGAGGCGCGGAGATCTGGCGGGTCTGGGTTCGGTTCGGTTCGGTTCGG 1560
QY 1561 GGGCGGTGTGGGCTCAGTGTGTCTGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGG 1620
Db 1561 GGGCGGTGTGGGCTCAGTGTGTCTGGTTCGGTTCGGTTCGGTTCGGTTCGGTTCGG 1620
QY 1621 ATTCTCGGAGCGCGGATATCGCGCGCGCGCTTGTGCGAGGACCGGTTCGCGACT 1680
Db 1621 ATTCTCGGAGCGCGGATATCGCGCGCGCGCTTGTGCGAGGACCGGTTCGCGACT 1680
QY 1681 TCCCGCGCTCGCCCTCGACCGCTCGCATGTTCGCCAAGTGGGGCCACAGGTGTCTCA 1740

Db 1681 TCCCGCGCTCGCCCTCGACCGCTCGCATGTTCGCCAAGTGGGGCCACAGGTGTCTCA 1740
QY 1741 ACATCAACACCAACTGGGCTACAAACCGCTGGGCGCGGAGCCGATCGTATCG 1800
Db 1741 ACATCAACACCAACTGGGCTACAAACCGCTGGGCGCGGAGCCGATCGTATCG 1800
QY 1801 ATCCCAACGGTGTCTGACCAACCAACCGATGATCGGGCGCGCACCGATCAATG 1860
Db 1801 ATCCCAACGGTGTCTGACCAACCAACCGATGATCGGGCGCGCACCGATCAATG 1860
QY 1861 CGTTCAAGCTCGGCTCGGCGCAAAACCTAGCGGTCTGATGTGTGCGGTATGACCGCACCC 1920
Db 1861 CGTTCAAGCTCGGCTCGGCGCAAAACCTAGCGGTCTGATGTGTGCGGTATGACCGCACCC 1920
QY 1921 AGGATGTGCGGCTGCTGACGCTGCGGTGCGCGCTGCGGTGCGGTGCGGTGCGGTG 1980
Db 1921 AGGATGTGCGGCTGCTGACGCTGCGGTGCGCGCTGCGGTGCGGTGCGGTGCGGTG 1980
QY 1981 GCGGCGTTCGGGTGAGCGCTGTCGCGCATGCGCAACAGCGGTGGGCGAGCGGAA 2040
Db 1981 GCGGCGTTCGGGTGAGCGCTGTCGCGCATGCGCAACAGCGGTGGGCGAGCGGAA 2040
QY 2041 GCGGCGTTCGGGTGAGCGCTGTCGCGCATGCGCGCTCGGCGCAACCGGTGCGGATT 2100
Db 2041 GCGGCGTTCGGGTGAGCGCTGTCGCGCATGCGCGCTCGGCGCAACCGGTGCGGATT 2100
QY 2101 CGCTGACCGGTGCGGAGAGACATTTAAACGGGTTCAGTTCAGTTCGATCCGATCCAGC 2160
Db 2101 CGCTGACCGGTGCGGAGAGACATTTAAACGGGTTCAGTTCAGTTCGATCCGATCCAGC 2160
QY 2161 CCGGTGATTCGGGCGGCGCTCGTCAACCGGCTTAGGACAGGTGCGGTATGAACACGG 2220
Db 2161 CCGGTGATTCGGGCGGCGCTCGTCAACCGGCTTAGGACAGGTGCGGTATGAACACGG 2220
QY 2221 CCGGTCTTAGGATATCCATCAGCTGCGGCGCTCGGCGATCCGNTGTAACAAAG 2280
Db 2221 CCGGTCTTAGGATATCCATCAGCTGCGGCGCTCGGCGATCCGNTGTAACAAAG 2280
QY 2281 CCGGAAA 2287
Db 2281 CCGGAAA 2287

RESULT 4
BD251333
LOCUS
DEFINITION Fused protein of Mycobacterium tuberculosis antigen and utilization thereof
ACCESSION BD251333
VERSION BD251333.1 GI:33061103
KEYWORDS JP 2002510494-A/12.
SOURCE synthetic construct
ORGANISM artificial sequences.
REFERENCE 1 (bases 1 to 1797)
AUTHORS Skeiky, Y.A.W., Alderson, M. and Neto, A.C.
TITLE Fused protein of Mycobacterium tuberculosis antigen and utilization thereof
JOURNAL Patent: JP 2002510494-A 12 09-APR-2002;
CORIXA CORP
OS Artificial Sequence
PN JP 2002510494-A/12
PD 09-APR-2002
PF 07-APR-1999 JP 2000542460
PR 07-APR-1998 US 09/056556, 30-DEC-1998 US 09/223040 PI
YASIR A W SKEIKY, MARK ALDERSON, ANTONIO CAMPOS NETO PC
C12N15/09, A61K39/04, A61K48/00, A61P31/04, C07K14/35, C07K19/00, PC
C12P21/02
PC C12N15/00
CC Description of Artificial Sequence: bi-fusion
protein TbH9-Ra35
CC (designated)
CC Mtbs9f)

[illegible]

TITLE		Fusion proteins of Mycobacterium tuberculosis antigens and their uses									
JOURNAL		Patent: US 6627198-A 25 30-SEP-2003;									
FEATURES		Location/Qualifiers									
source		1..1797									
		/organism="unknown"									
ORIGIN		/mol_type="genomic DNA"									
Query Match		77.4%; Score 1770.8; DB 6; Length 1797;									
Best Local Similarity		99.9%; Pred. No. 1.7e-209;									
Matches 1772; Conservative		0; Mismatches 2; Indels 0; Gaps 0;									
QY	464	CATGTTGATTTTCGGGGCGTTACCAACCGAGATCAACTCCGCGAGGATGATCGCGGCC	523								
DB	24	CATGTTGATTTTCGGGGCGTTACCAACCGAGATCAACTCCGCGAGGATGATCGCGGCC	83								
QY	524	GGGTTCCGGCTCGCTGGTGGCGCGCTCAGATGTTGGACAGCGTGGCGAGTGAACCTTT	583								
DB	84	GGGTTCCGGCTCGCTGGTGGCGCGCTCAGATGTTGGACAGCGTGGCGAGTGAACCTTT	143								
QY	584	TTCCGGCGCGTTCGAGTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT	643								
DB	144	TTCCGGCGCGTTCGAGTTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGAT	203								
QY	644	TTCCGTCCGGCGCTTCATGTTGGCGCGCTTCGCGCTATGTTGGCGTGGATGAGCGTCA	703								
DB	204	TTCCGTCCGGCGCTTCATGTTGGCGCGCTTCGCGCTATGTTGGCGTGGATGAGCGTCA	263								
QY	704	CGCGGGCAGGCGAGCTGACCGCGCGCTCAGTTCGCGCTTCGCGCTTCGCGCTTCG	763								
DB	264	CGCGGGCAGGCGAGCTGACCGCGCGCTCAGTTCGCGCTTCGCGCTTCGCGCTTCG	323								
QY	764	GGCGTATGGCTGACGTTCCCGCGCTGATTCGCGAGAACCTGCTGTAACCTGATGAT	823								
DB	324	GGCGTATGGCTGACGTTCCCGCGCTGATTCGCGAGAACCTGCTGTAACCTGATGAT	383								
QY	824	TCTGATAGCGACCACTCTTCGGGCAAAACCCCGCGATTCGCGCTCAACGAGCGCGA	883								
DB	384	TCTGATAGCGACCACTCTTCGGGCAAAACCCCGCGATTCGCGCTCAACGAGCGCGA	443								
QY	884	ATACGCGAGATGTGGGCGCAAGACCGCGCGATGTTGCTACGCGCGCGCGCGCG	943								
DB	444	ATACGCGAGATGTGGGCGCAAGACCGCGCGATGTTGCTACGCGCGCGCGCGCG	503								
QY	944	GACGCGAGCGCGACGTTGCTGCGCTTCGAGGAGCGCGCGAGATGACGAGCGCGGTG	1003								
DB	504	GACGCGAGCGCGACGTTGCTGCGCTTCGAGGAGCGCGCGAGATGACGAGCGCGGTG	563								
QY	1004	GCTCTCGAGAGCGCGCGCTGAGGAGCGCTCCGACACCGCGCGCGCGCGCGCG	1063								
DB	564	GCTCTCGAGAGCGCGCGCTGAGGAGCGCTCCGACACCGCGCGCGCGCGCGCG	623								
QY	1064	GATGACAAATGTCCCGAGCGCTGCAACAGCTGGCGCGCGCGCGCGCGCGCGCG	1123								
DB	624	GATGACAAATGTCCCGAGCGCTGCAACAGCTGGCGCGCGCGCGCGCGCGCGCG	683								
QY	1124	TTCTTCCAAAGCTGGTGGCTGGAGAGCGTTCCTCCCGCATTCGCTCCCGCATCAG	1183								
DB	684	TTCTTCCAAAGCTGGTGGAGAGCGTTCCTCCCGCATTCGCTCCCGCATCAGCAA	743								
QY	1184	CATGTTGTCATGCCCAACCAACATGTCATGACCAACTCGGTTGTCATGACCAA	1243								
DB	744	CATGTTGTCATGCCCAACCAACATGTCATGACCAACTCGGTTGTCATGACCAA	803								
QY	1244	CACCTTGAGCTCGATGTTGAAGGCTTTCCTCCGGCGCGCGCGCGCGCGCGCG	1303								
DB	804	CACCTTGAGCTCGATGTTGAAGGCTTTCCTCCGGCGCGCGCGCGCGCGCGCG	863								
QY	1304	CGCGCGCAAAACCGGGTTCGGGCGATGAGCTCGCTGGGCGAGCTCGCTGGTTC	1363								
DB	864	CGCGCGCAAAACCGGGTTCGGGCGATGAGCTCGCTGGGCGAGCTCGCTGGTTC	923								

QY	1364	TTCTGGCGGTGGGTGGCGCGCAACTTGGGTGGGGCGGCTCGGTGGTTGTTGCGT	1423								
DB	924	TTCTGGCGGTGGGTGGCGCGCAACTTGGGTGGGGCGGCTCGGTGGTTGTTGCGT	983								
QY	1424	GCCGAGGCTGGCGCGCGCGCAACACGAGTACACCCGGCGCGCGCGCGCGCGT	1483								
DB	984	GCCGAGGCTGGCGCGCGCGCAACACGAGTACACCCGGCGCGCGCGCGCGCGT	1043								
QY	1484	GACAGCTGACAGCGCGCGGAGAGAGGCGCGCGGAGATGCTGGCGGGGCTGCGGT	1543								
DB	1044	GACAGCTGACAGCGCGCGGAGAGAGGCGCGGAGATGCTGGCGGGGCTGCGGT	1103								
QY	1544	GGGCGAGATGGGCGCGCGCGGTGGTGGTTCAGTGGTGTCTGCTGCTGCTGCTG	1603								
DB	1104	GGGCGAGATGGGCGCGCGCGGTGGTGGTTCAGTGGTGTCTGCTGCTGCTGCTG	1163								
QY	1604	ACCTATGATGCGCGATTTCTCGGAGCGCGGATATCGCCCGCGCGCGCTTGTGCA	1663								
DB	1164	ACCTATGATGCGCGATTTCTCGGAGCGCGGATATCGCCCGCGCGCGCTTGTGCA	1223								
QY	1664	GGACCGGTTCGCGGACTTTCGCGCGCTCGACCGCTCGCGGATGCTGCGCGCAAGT	1723								
DB	1224	GGACCGGTTCGCGGACTTTCGCGCGCTCGACCGCTCGCGGATGCTGCGCGCAAGT	1283								
QY	1724	GGGCGCAGAGTGGTCAACATCAACACCACTGGGCTACAAACGCGTGGCGCGCG	1783								
DB	1284	GGGCGCAGAGTGGTCAACATCAACACCACTGGGCTACAAACGCGTGGCGCGCG	1343								
QY	1784	GACCGGATCGTCATCGATCCCAACGCTGCTGCTGACCAACCAACACGCTGATG	1843								
DB	1344	GACCGGATCGTCATCGATCCCAACGCTGCTGCTGACCAACCAACACGCTGATG	1403								
QY	1844	GGCACCGCATCAATCGTTCAAGCTCGGCTCGCGCAAAACCTACGCGCTGATG	1903								
DB	1404	GGCACCGCATCAATCGTTCAAGCTCGGCTCGCGCAAAACCTACGCGCTGATG	1463								
QY	1904	CGGATGATGCGGACCGCGCGGATGTCGCGGTGCTGACGCTCGCGGTGCGCTGCC	1963								
DB	1464	CGGATGATGCGGACCGCGCGGATGTCGCGGTGCTGACGCTCGCGGTGCGCTGCC	1523								
QY	1964	GTCGCGCGCATCGGTGGCGCGCTGCGGTGCTGAGCCCTGCTCGCATGGGCAACAG	2023								
DB	1524	GTCGCGCGCATCGGTGGCGCGCTGCGGTGCTGAGCCCTGCTCGCATGGGCAACAG	1583								
QY	2024	CGGTGGCGCGGCGGACCGCGCGGCTGCGGTGCTGCGGAGGCTGCTCGCGTGG	2083								
DB	1584	CGGTGGCGCGGCGGACCGCGCGGCTGCGGTGCTGCGGAGGCTGCTCGCGTGG	1643								
QY	2084	CGTCAGCGCTCGGATTCGCTGACCGGTGCGGAGAGACATTGAACCGGTTGAT	2143								
DB	1644	CGTCAGCGCTCGGATTCGCTGACCGGTGCGGAGAGACATTGAACCGGTTGAT	1703								
QY	2144	CGATGCGCGCATTCAGCGCGGTGATTCGGCGCGGCGCTGCTCAACGGCGCTAG	2203								
DB	1704	CGATGCGCGCATTCAGCGCGGTGATTCGGCGCGGCGCTGCTCAACGGCGCTAG	1763								
QY	2204	GGTCGGTATGAACACGCGCGCTCTAGGATATC	2237								
DB	1764	GGTCGGTATGAACACGCGCGCTCTAGGATATC	1797								

RESULT 6
BD251331
LOCUS BD251331
DEFINITION Fused protein of Mycobacterium tuberculosis antigen and utilization thereof.
ACCESSION BD251331
VERSION BD251331.1
KEYWORDS JP 2002510494-A/10.
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 1801)

1801 bp DNA linear PAT 17-JUL-2003

144 TTCTGCGCGGCTCGGGGTTTCAGTCGGTGTCTGGGGTCTGACGGTGGGTCTGTGATAGG 203
644 TTCTGTCGCGGCTCTGATGTCGCGCGGCTTCGCCGTATGTCGGTGTGATGAGCGTCAAC 703
204 TTCTGTCGCGGCTCTGATGTCGCGCGGCTTCGCCGTATGTCGGTGTGATGAGCGTCAAC 263
704 CCGCGGCGAGCGGCTGACCGCGCGGCTTCGCCGTATGTCGGTGTGATGAGCGTCAAC 763
264 CCGCGGCGAGCGGCTGACCGCGCGGCTTCGCCGTATGTCGGTGTGATGAGCGTCAAC 323
764 GCGGTATGCGGCTGACCGGCTTCGCCGTATGTCGGTGTGATGAGCGTCAAC 823
324 GCGGTATGCGGCTGACCGGCTTCGCCGTATGTCGGTGTGATGAGCGTCAAC 383
824 TCTGATAGCGACCACTTCTGGGCGAACAACCGCGGCTTCGCCGTATGAGCGGCTCAAC 883
384 TCTGATAGCGACCACTTCTGGGCGAACAACCGCGGCTTCGCCGTATGAGCGGCTCAAC 443
884 ATACGGCGAGATGTGGGCGCAAGACCGCGCGGCTTCGCCGTATGAGCGGCTCAAC 943
444 ATACGGCGAGATGTGGGCGCAAGACCGCGCGGCTTCGCCGTATGAGCGGCTCAAC 503
944 GACGGCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1003
504 GACGGCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 563
1004 GCTCTCGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1063
564 GCTCTCGAGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 623
1064 GATGAACAATGTGCGCGCGGCTGCAACAGTGTGCGGCGGCTGCAACAGTGTGCGGCGG 1123
624 GATGAACAATGTGCGCGCGGCTGCAACAGTGTGCGGCGGCTGCAACAGTGTGCGGCGG 683
1124 TTCTTCCAGCTGGGTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1183
684 TTCTTCCAGCTGGGTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 743
1184 CATGTGTGATGCGGCAACACATGTCGATGACCACTCGGCTGCTGCTGCTGCTGCTGCTGCT 1243
744 CATGTGTGATGCGGCAACACATGTCGATGACCACTCGGCTGCTGCTGCTGCTGCTGCTGCT 803
1244 CACTTGTAGCTGATGTTGAAGGCTTGTCTCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 1303
804 CACTTGTAGCTGATGTTGAAGGCTTGTCTCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCT 863
1304 CCGCGGCGAACAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1363
864 CCGCGGCGAACAACCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 923
1364 TCTGGCGGTGGGTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1423
924 TCTGGCGGTGGGTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 983
1424 GCGCGAGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1483
984 GCGCGAGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1043
1484 GACGAGCTGACGAGCGGCGGCAAGAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1543
1044 GACGAGCTGACGAGCGGCGGCAAGAGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1103
1544 GCGCGAGATGGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1603
1104 GCGCGAGATGGGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1163
1604 ACCCTATGATGCGGCTTCTCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1655
1164 ACCCTATGATGCGGCTTCTCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1215

RESULT 8

BD273807
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
source
ORIGIN

Query Match 51.1%; Score 1169; DB 6; Length 1188;
Best Local Similarity 99.6%; Pred. No. 3.4e-135;
Matches 1172; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 463 TCATGTTGGATTTCGGGGCGTTACACCGGAGATCAACTCCGCGAGGATGTACCGCGGCC 522
Db 11 TAAATGTTGGATTTCGGGGCGTTACACCGGAGATCAACTCCGCGAGGATGTACCGCGGCC 70
QY 523 CGGTTTCGGCTTCGCTGGTGGCGCGGCTCAGATGGGACAGCGTGGCGAGTGAACCTGT 582
Db 71 CGGTTTCGGCTTCGCTGGTGGCGCGGCTCAGATGGGACAGCGTGGCGAGTGAACCTGT 130
QY 583 TTTTCGGCGCGTTCGCGGTTTCAGTTCGGTGGTTCAGCGGTCGACGTTGGGGTCGTTGGGATAG 642
Db 131 TTTTCGGCGCGTTCGCGGTTTCAGTTCGGTGGTTCAGCGGTCGACGTTGGGGTCGTTGGGATAG 190
QY 643 GTTTCGTCGGCGGCTCAGATGGTGGCGCGGCTTCGCGGATATGTGCGGTGGATGAGCGTCA 702
Db 191 GTTTCGTCGGCGGCTCAGATGGTGGCGCGGCTTCGCGGATATGTGCGGTGGATGAGCGTCA 250
QY 703 CCGCGGCGAGCGGCGGCTGACCGCGGCGGCGGCTGCGGCTGCGGCTGCGGCGGCTTACGAGA 762
Db 251 CCGCGGCGAGCGGCGGCTGACCGCGGCGGCGGCTGCGGCTGCGGCTGCGGCGGCTTACGAGA 310
QY 763 CCGCGTATGGCTGACCGGTCGCGGCGGCGGCTGATCGCGGAGAACCGTGTGAACCTGATGA 822
Db 311 CCGCGTATGGCTGACCGGTCGCGGCGGCGGCTGATCGCGGAGAACCGTGTGAACCTGATGA 370
QY 823 TTCTGATAGGACCAACCTCTTGGGGCAAAACACCGCGGCGATCGCGTCAACGAGCGCG 882

371	Db	TTCTGATAGCGACCAACCTCTTTGGGGCAAAACACCCCGGGGATCGCGTTCACAGAGCCG	430
883	Qy	AATACGGCGAGATGTGGGCCCAAGACGCGCGCGATGTTTGGCTACGCCCGCGCGACGG	942
431	Db	AATACGGCGAGATGTGGGCCCAAGACGCGCGCGATGTTTGGCTACGCCCGCGCGACGG	490
943	Qy	CGACGGGACGGCGACGTTGCTCGGTTTCAGAGAGGCGCGGAGATGACACAGCCCGGGT	1002
491	Db	CGACGGGACGGCGACGTTGCTCGGTTTCAGAGAGGCGCGGAGATGACACAGCCCGGGT	550
1003	Qy	GGCTTCCTCGAGCAGCGCGCGTTCAGAGAGGCTCCGACACCGCGCGCGCAACCACT	1062
551	Db	GGCTTCCTCGAGCAGCGCGCGTTCAGAGAGGCTCCGACACCGCGCGCGCAACCACT	610
1063	Qy	TGATGAACAATGTGCCCCAGCGCTGCAACAGCTGGCCAGCCACACAGGCGACCAACGC	1122
611	Db	TGATGAACAATGTGCCCCAGCGCTGCAACAGCTGGCCAGCCACACAGGCGACCAACGC	670
1123	Qy	CTTCTTCAGACTGGGTGGCTGTGGAAGACGGTCTCGCCGATCGGTTCGCCGATCAGCA	1182
671	Db	CTTCTTCAGACTGGGTGGCTGTGGAAGACGGTCTCGCCGATCGGTTCGCCGATCAGCA	730
1183	Qy	ACATGGTGTTCGATGGCCAAACACACATGTCGATGACCAACTCGGGTGTTCGATGACCA	1242
731	Db	ACATGGTGTTCGATGGCCAAACACACATGTCGATGACCAACTCGGGTGTTCGATGACCA	790
1243	Qy	ACACCTTGAGCTCGATGTTGAAGGGCTTTGCTCCGGGGGGCCGCCAGCCCGTGCAAA	1302
791	Db	ACACCTTGAGCTCGATGTTGAAGGGCTTTGCTCCGGGGGGCCGCCAGCCCGTGCAAA	850
1303	Qy	CCGGCGCGCAAAACGGGGTCCGGCGATGAGCTCGCTGGCGAGCTCGCTGGGTTCCTCG	1362
851	Db	CCGGCGCGCAAAACGGGGTCCGGCGATGAGCTCGCTGGCGAGCTCGCTGGGTTCCTCG	910
1363	Qy	GTCTGGCGCGTGGGGTGGCCCAACTTGGGTCCGGCGCGCTCGGTCCGTTCTGTTTCGG	1422
911	Db	GTCTGGCGCGTGGGGTGGCCCAACTTGGGTCCGGCGCGCTCGGTCCGTTCTGTTTCGG	970
1423	Qy	TGCCGACGGCTCGGGCGCGCCAAACAGACGATCAACCCCGGGCGCGCGCTGCCTC	1482
971	Db	TGCCGACGGCTCGGGCGCGGCCAAACAGACGATCAACCCCGGGCGCGCGCTGCCTC	1030
1483	Qy	TGACCGACTGACACGCGCGCGGAAGAGGGCCCGGGCAGATGCTGGCGGGCTGCCCG	1542
1031	Db	TGACCGACTGACACGCGCGCGGAAGAGGGCCCGGGCAGATGCTGGCGGGCTGCCCG	1090
1543	Qy	TGGGGCAGATGGGGCGCCAGGGCGGGTGGTGGGCTCATGTGTGTCTCGCTGTTCGGCCG	1602
1091	Db	TGGGGCAGATGGGGCGCCAGGGCGGGTGGTGGGCTCATGTGTGTCTCGCTGTTCGGCCG	1150
1603	Qy	GACCCTATGTATGCGCATTTCCGGCAGCCGGCGA	1639
1151	Db	GACCCTATGTATGCGCATTTCCGGCAGCCGGCTA	1187

RESULT 9	
AX004989	
LOCUS	1188 bp DNA linear PAT 24-AUG-2000
DEFINITION	Sequence 108 from Patent WO9909186.
AX004989	
ACCESSION	
AX004989.1	GI:9938350
KEYWORDS	
SOURCE	Mycobacterium tuberculosis
ORGANISM	Mycobacterium tuberculosis
	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
	Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
	tuberculosis complex.
REFERENCE	1
AUTHORS	Portnoi,D. and Guigueno,A.
TITLE	Polypeptide nucleic sequences exported from mycobacteria, vectors
	comprising same and uses for diagnosing and preventing tuberculosis
JOURNAL	Patent : WO 9909186-A 108 25-FEB-1999;
	PORTNOI DENIS (FR); GUIGUENO AGNES (FR)


```

VERSION      AR182495.1  GI:20225702
KEYWORDS     Unknown.
SOURCE       Unknown.
ORGANISM     Unclassified.
REFERENCE    1 (bases 1 to 3058)
AUTHORS      Reed, S.G., Skeiky, Y.A.W., Dillon, D.C., Campos-Neto, A., Houghton, R.,
              Vedvick, R.S., and Twardzik, D.R.
TITLE        Compounds and methods for diagnosis of tuberculosis
JOURNAL      Patent: US 638952-A 101 15-JAN-2002;
FEATURES     Location/Qualifiers
              source
              1..3058
              /organism="unknown"
              /mol_type="unassigned DNA"

ORIGIN
Query Match      51.1%; Score 1169; DB 6; Length 3058;
Best Local Similarity 99.6%; Pred. No. 2.7e-135;
Matches 1172; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 463 TCATGGTGGATTTCGGGGCGTTACACCGGAGATCAACTCCGGAGGATGTAGCCGGCC 522
DB 391 TAATGGTGGATTTCGGGGCGTTACACCGGAGATCAACTCCGGAGGATGTAGCCGGCC 450
QY 523 CGGGTTCGGCTCGCTGGTGGCCCGGCTCAGATGTGGGACAGCGTGGCCAGTGAACCTGT 582
DB 451 CGGGTTCGGCTCGCTGGTGGCCCGGCTCAGATGTGGGACAGCGTGGCCAGTGAACCTGT 510
QY 583 TTTCCGGCGCGTCCGGCTTCAGTTCGGTGTCTGGGTCTGACGGTGGGTCTGGATAG 642
DB 511 TTTCCGGCGCGTCCGGCTTCAGTTCGGTGTCTGGGTCTGACGGTGGGTCTGGATAG 570
QY 643 GTTCGTTCGGCGGCTCATGTGTGGCGCGCTCCGCGTATGTGGGTGATGATGAGGTCA 702
DB 571 GTTCGTTCGGCGGCTCATGTGTGGCGCGCTCCGCGTATGTGGGTGATGAGGTCA 630
QY 703 CCGCGGCGAGCGGAGCTGACGCCCGCCAGGTCGGGTTCGGTGGCGGCTTACGAGA 762
DB 631 CCGCGGCGAGCGGAGCTGACGCCCGCCAGGTCGGGTTCGGTGGCGGCTTACGAGA 690
QY 763 CGGGTATGGCTACCGTCCCGCGCGTGTGCGCGAGACCGTGTCTGACTGATCA 822
DB 691 CGGGTATGGCTACCGTCCCGCGCGTGTGCGCGAGACCGTGTCTGACTGATCA 750
QY 823 TTTCTAGTAGCAACCACTCTTGGGGCAAAACACCCCGGCGATCGGGTCAACGAGGCGG 882
DB 751 TTTCTAGTAGCAACCACTCTTGGGGCAAAACACCCCGGCGATCGGGTCAACGAGGCGG 810
QY 883 AATACGGCGAGATGTGGGCGCAAGACCGCGCGGATGTTGGCTAGCGCGCGGACGG 942
DB 811 AATACGGCGAGATGTGGGCGCAAGACCGCGCGGATGTTGGCTAGCGCGCGGACGG 870
QY 943 CGACCGCGAGCGGACGCTTCTGCGGTTCCAGAGGCGCGGAGATGACACGCGCGGGTG 1002
DB 871 CGACCGCGAGCGGACGCTTCTGCGGTTCCAGAGGCGCGGAGATGACACGCGCGGGTG 930
QY 1003 GGCTCTTCGAGAGCGCGCGGTCGAGAGGCTCCGACACCGCGCGGCGAACAGT 1062
DB 931 GGCTCTTCGAGAGCGCGCGGTCGAGAGGCTCCGACACCGCGCGGCGAACAGT 990
QY 1063 TGATGAACAATGTGCGCGAGCGCTGCAACAGTGGCGGCGGCGGCGGCGGCGGCGG 1122
DB 991 TGATGAACAATGTGCGCGAGCGCTGCAACAGTGGCGGCGGCGGCGGCGGCGGCGG 1050
QY 1123 CTTCTTCCAAAGCTGGTGGCTGTGGAAGACGGTCTCGCGCATCGGTCCCGATCAGCA 1182
DB 1051 CTTCTTCCAAAGCTGGTGGCTGTGGAAGACGGTCTCGCGCATCGGTCCCGATCAGCA 1110
QY 1183 ACATGGTGTGATGGCCCAACCAACATGTCATGACCACTCGGGTGTGTCATGACCA 1242
DB 1111 ACATGGTGTGATGGCCCAACCAACATGTCATGACCACTCGGGTGTGTCATGACCA 1170
QY 1243 ACACCTTGAGCTGATTTGAAGGGCTTTTGCTCCGGCGGCGGCGGCGGCGGCGTGC 1302
```

```

DB 1171 ACACCTTGAGCTGATTTGAAGGGCTTTTGCTCCGGCGGCGGCGGCGGCGTGC 1230
QY 1303 CGCGCGCGCAAAACCGGGGTCCGGGCGATGAGTCTGCTGGGAGAGTCTGCTGGGTCTTTCGG 1362
DB 1231 CGCGCGCGCAAAACCGGGGTCCGGGCGATGAGTCTGCTGGGAGAGTCTGCTGGGTCTTTCGG 1290
QY 1363 GTCTGGGCGGTGGGTGGCGGCAACTTGGGTGGGCGGCGCTCGTGGTTCGTTTCGG 1422
DB 1291 GTCTGGGCGGTGGGTGGCGGCAACTTGGGTGGGCGGCGCTCGTGGTTCGTTTCGG 1350
QY 1423 TGCCGCGAGGCTGGGCGCGGCGCAACAGGAGTCAACCCCGCGGCGGCGGCGGCTGCGCGC 1482
DB 1351 TGCCGCGAGGCTGGGCGCGGCGCAACAGGAGTCAACCCCGCGGCGGCGGCGGCTGCGCGC 1410
QY 1483 TGACCAAGCTGACAGCGCGCGGAAAGAGGCGCGGCGAGATGCTGGGCGGCGGCTGCGCGC 1542
DB 1411 TGACCAAGCTGACAGCGCGCGGAAAGAGGCGCGGCGAGATGCTGGGCGGCGGCTGCGCGC 1470
QY 1543 TGGGCGAGATGGCGCGGCGGCGGCTGGTGGCTCAGTGGTGTCTGCTGGTTCGCGCGC 1602
DB 1471 TGGGCGAGATGGCGCGGCGGCGGCTGGTGGCTCAGTGGTGTCTGCTGGTTCGCGCGC 1530
QY 1603 GACCCATGTGATGCGCGCAATTCCTCGGCGAGCGCGGA 1639
DB 1531 GACCCATGTGATGCGCGCAATTCCTCGGCGGCGCGGCTA 1567

RESULT 12
AR194878
LOCUS       AR194878
DEFINITION Sequence 106 from patent US 6350456.
ACCESSION  AR194878
VERSION    AR194878.1  GI:20244315
KEYWORDS   Unknown.
SOURCE     Unknown.
ORGANISM   Unknown.
REFERENCE  1 (bases 1 to 3058)
AUTHORS    Reed, S.G., Skeiky, Y.A.W. and Dillon, D.C.
TITLE      Compositions and methods for the prevention and treatment of M.
            tuberculosis infection
JOURNAL    Patent: US 6350456-A 106 26-FEB-2002;
FEATURES   Location/Qualifiers
            source
            1..3058
            /organism="unknown"
            /mol_type="unassigned DNA"

ORIGIN
Query Match      51.1%; Score 1169; DB 6; Length 3058;
Best Local Similarity 99.6%; Pred. No. 2.7e-135;
Matches 1172; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 463 TCATGGTGGATTTCGGGGCGTTACACCGGAGATCAACTCCGGAGGATGTAGCCGGCC 522
DB 391 TAATGGTGGATTTCGGGGCGTTACACCGGAGATCAACTCCGGAGGATGTAGCCGGCC 450
QY 523 CGGGTTCGGCTCGCTGGTGGCCCGGCTCAGATGTGGGACAGCGTGGCCAGTGAACCTGT 582
DB 451 CGGGTTCGGCTCGCTGGTGGCCCGGCTCAGATGTGGGACAGCGTGGCCAGTGAACCTGT 510
QY 583 TTTCCGGCGCGTCCGGCTTCAGTTCGGTGTCTGGGTCTGACGGTGGGTCTGGATAG 642
DB 511 TTTCCGGCGCGTCCGGCTTCAGTTCGGTGTCTGGGTCTGACGGTGGGTCTGGATAG 570
QY 643 GTTCGTTCGGCGGCTCATGTGTGGCGCGCTCCGCGTATGTGGGTGATGATGAGGTCA 702
DB 571 GTTCGTTCGGCGGCTCATGTGTGGCGCGCTCCGCGTATGTGGGTGATGAGGTCA 630
QY 703 CCGCGGCGAGCGGAGCTGACGCCCGCCAGGTCGGGTTCGGTGGCGGCTTACGAGA 762
DB 631 CCGCGGCGAGCGGAGCTGACGCCCGCCAGGTCGGGTTCGGTGGCGGCTTACGAGA 690
```

763 CGCGCTATGGGTGACGGTGCCTCCCGCGGTGATCGCCGAGAACCGTCTGACATGATGA 822
Db CGCGCTATGGGTGACGGTGCCTCCCGCGGTGATCGCCGAGAACCGTCTGACATGATGA 750
QY TTCTGATACGACCAACCTCTTTGGGGCAAAAACACCCCGCGGATCGCGGTCAACGAGGCGG 882
Db TTCTGATACGACCAACCTCTTTGGGGCAAAAACACCCCGCGGATCGCGGTCAACGAGGCGG 810
QY AATACGGCAGATGTGGGCGCCAAAGACCGCCCGCGCGATTTGGCTACCGCGCGGCGAGG 942
Db AATACGGCAGATGTGGGCGCCAAAGACCGCCCGCGCGATTTGGCTACCGCGCGGCGAGG 870
QY CGACGGCAGCGGACGCTTTGGGGCAAAAACACCCCGCGGATCGCGGTCAACGAGGCGG 1002
Db CGACGGCAGCGGACGCTTTGGGGCAAAAACACCCCGCGGATCGCGGTCAACGAGGCGG 930
QY GGCTCTCCAGCAGGCGCGCGGTGCGGTGCGAGGCGCGCGGATGACACGCGCGGCGG 1062
Db GGCTCTCCAGCAGGCGCGCGGTGCGAGGCGCGCGGTGCGAGGCGCGCGGTGCGAGGCGG 990
QY TGATGAACATGTGCGGCGCGGTGCGAGGCGCGCGGTGCGAGGCGCGCGGTGCGAGGCGG 1122
Db TGATGAACATGTGCGGCGCGGTGCGAGGCGCGCGGTGCGAGGCGCGCGGTGCGAGGCGG 1050
QY CTTCTTCCAGCTGGGTGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1182
Db CTTCTTCCAGCTGGGTGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1110
QY ACATGCTGATGCGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1242
Db ACATGCTGATGCGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1170
QY ACACCTTTGAGCTGATGTTGAGGGGCTTTGCTCCGGCGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1302
Db ACACCTTTGAGCTGATGTTGAGGGGCTTTGCTCCGGCGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1230
QY CGCGCGCGCAAAACGGGGTCCGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1362
Db CGCGCGCGCAAAACGGGGTCCGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1290
QY GTCTGGCGCGGTGGGGTGGCGCGCACTTGGGTGGCGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1422
Db GTCTGGCGCGGTGGGGTGGCGCGCACTTGGGTGGCGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1350
QY TGCGCGAGGCGCTGGCGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1482
Db TGCGCGAGGCGCTGGCGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1410
QY TGACCAAGCTGACCGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1542
Db TGACCAAGCTGACCGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1470
QY TGGGGCAGATGGCGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1602
Db TGGGGCAGATGGCGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1530
QY GACCTATGATGCGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1639
Db GACCTATGATGCGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1567

RESULT 13

AR233150
LOCUS AR233150 3058 bp DNA linear PAT 20-DEC-2002
DEFINITION Sequence 101 from patent US 6458366.
ACCESSION AR233150
VERSION AR233150.1 GI:27275586
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 3058)
AUTHORS Reed,S.G., Skeiky,Y.A.W., Dillon,D.C., Campos-Neto,A., Houghton,R.,

Vedvick,T.S., Twardzik,D.R., Lodes,M.J. and Hendrickson,R.C.
Compounds and methods for diagnosis of tuberculosis
Patent: US 6458366-A 101 01-OCI-2002;
Location/Qualifiers
1. 3058
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Query Match 51.1%; Score 1169; DB 6; Length 3058;
Best Local Similarity 99.6%; Pred. No. 2.7e-135;
Matches 1172; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 463 TCATGTTGGGTTTCGGGGCGTTACACCGGAGATCAACTCCGGGAGGATGTAAGCCGCGCC 522
Db 391 TAATGTTGGGTTTCGGGGCGTTACACCGGAGATCAACTCCGGGAGGATGTAAGCCGCGCC 450
QY 523 GGGTTTCGGGCTCGCTGGTGGCGCGCTCAGATGTGGGACAGGTGGCGAGTGAACCTGT 582
Db 451 CGGTTTCGGGCTCGCTGGTGGCGCGCTCAGATGTGGGACAGGTGGCGAGTGAACCTGT 510
QY 583 TTTCCGGCGCGGTTCAGTTCGGTGTCTGGGGTCTGAGGTGGGTCTGTTGGATAG 642
Db 511 TTTCCGGCGCGGTTCAGTTCGGTGTCTGGGGTCTGAGGTGGGTCTGTTGGATAG 570
QY 643 GTTCTGCGGGGTCTGATGTGGCGCGCTTCCCGTATGTGGCTGATGATGACGTCA 702
Db 571 GTTCTGCGGGGTCTGATGTGGCGCGCTTCCCGTATGTGGCTGATGATGACGTCA 630
QY 703 CCGCGGGCAGGCGGAGCTGACCGCGCGCGGTTCGGGTTCGCGCGCGCTTACGAGA 762
Db 631 CCGCGGGCAGGCGGAGCTGACCGCGCGCGGTTCGGGTTCGCGCGCGCTTACGAGA 690
QY 763 CGGCTATGGGTGACCGTGCCTCCCGCGGTGATCGCGGAGAACCGTGTGTAATGATGA 822
Db 691 CGGCTATGGGTGACCGTGCCTCCCGCGGTGATCGCGGAGAACCGTGTGTAATGATGA 750
QY 823 TTTCTAGTAGAACCAACTCTTTGGGGCAAAACACCCCGGAGATCGCGGTCAACGAGCGG 882
Db 751 TTTCTAGTAGAACCAACTCTTTGGGGCAAAACACCCCGGAGATCGCGGTCAACGAGCGG 810
QY 883 AATACGGCAGATGGGCGCGCAAGACCGCGCGGTGTTGGCTACCGCGCGCGAGG 942
Db 811 AATACGGCAGATGGGCGCGCAAGACCGCGCGGTGTTGGCTACCGCGCGCGAGG 870
QY 943 CGACGGCAGCGCGACGTTGCTGCTTCGAGGAGGCGCGGAGATGACACGCGCGGTG 1002
Db 871 CGACGGCAGCGCGACGTTGCTGCTTCGAGGAGGCGCGGAGATGACACGCGCGGTG 930
QY 1003 GGCTCTCGACAGCGCGCGGTTCGAGGAGGCGCTCGACACCGCGCGCGGAGACCACT 1062
Db 931 GGCTCTCGACAGCGCGCGGTTCGAGGAGGCGCTCGACACCGCGCGCGGAGACCACT 990
QY 1063 TGATGAACATGTGGCGCGCGGTGCAACACAGCTGGGCGCGCGCGCGCGCGCGCGG 1122
Db 991 TGATGAACATGTGGCGCGCGGTGCAACACAGCTGGGCGCGCGCGCGCGCGCGCGG 1050
QY 1123 CTTCTTCCAGCTGGGTGGCGCTGTGGAAGACGCTTCGCGCGATCGGTTCGCGGTGAGCA 1182
Db 1051 CTTCTTCCAGCTGGGTGGCGCTGTGGAAGACGCTTCGCGCGATCGGTTCGCGGTGAGCA 1110
QY 1243 ACACCTTGAAGCTGATGTTGAGGGGCTTTGCTCCGGCGCGCGCGCGCGCGCGCGG 1302
Db 1171 ACACCTTGAAGCTGATGTTGAGGGGCTTTGCTCCGGCGCGCGCGCGCGCGCGCGG 1230
QY 1303 CCGCGCGCAAAACGGGGTCCGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1362
Db 1231 CCGCGCGCAAAACGGGGTCCGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1290
QY 1363 GTCTGGCGCGGTGGGGTGGCGCGCACTTGGGTGGCGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1422
Db 1291 GTCTGGCGCGGTGGGGTGGCGCGCACTTGGGTGGCGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1350
QY 1423 TGCGCGAGGCGCTGGCGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1482
Db 1351 TGCGCGAGGCGCTGGCGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1410
QY 1483 TGACCAAGCTGACCGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1542
Db 1411 TGACCAAGCTGACCGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1470
QY 1543 TGGGGCAGATGGCGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1602
Db 1471 TGGGGCAGATGGCGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1530
QY 1603 GACCTATGATGCGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1639
Db 1531 GACCTATGATGCGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGCGGTGCGAGGCGG 1567

QY 1363 GTCTGGCGGTGGGTGGCCGCCAACTTGGTCCGGCGGCTCGGTGGTTCGTTGTCGG 1422
 Db 1291 GTCTGGCGGTGGGTGGCCGCCAACTTGGTCCGGCGGCTCGGTGGTTCGTTGTCGG 1350
 QY 1423 TGCCGACGCTGGCGCGCGCCCAACAGGAGTCACTCCCGCGCGCGCGCGCTGCCGC 1482
 Db 1351 TGCCGACGCTGGCGCGCGCCCAACAGGAGTCACTCCCGCGCGCGCGCGCTGCCGC 1410
 QY 1483 TGACACGCTGACACAGCCCGCGGAAAGAGGCGCGGAGATGCTGGCGCGCGTGCCTG 1542
 Db 1411 TGACACGCTGACACAGCCCGCGGAAAGAGGCGCGGAGATGCTGGCGCGCGTGCCTG 1470
 QY 1543 TGGGCGACATGGCGCCAGGCGCGGTGGTGGGCTCAGTGGTGGTTCGCGCGC 1602
 Db 1471 TGGGCGACATGGCGCCAGGCGCGGTGGTGGGCTCAGTGGTGGTTCGCGCGC 1530
 QY 1603 GACCTATGTGATGCGGCACTTCTCCGCGAGCGCGCA 1639
 Db 1531 GACCTATGTGATGCGGCACTTCTCCGCGAGCGCGCA 1567

RESULT 14
 AR353355
 LOCUS
 DEFINITION Sequence 106 from patent US 6592877.
 ACCESSION AR353355
 VERSION AR353355.1 GI:33759161
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 3058)
 AUTHORS Read, S.G., Skeiky, Y.A.W., Dillon, D.C., Campos-Neto, A., Houghton, R.,
 Vedvick, T.S., Twardzik, D.R., Lodes, M.J. and Hendrickson, R.C.
 TITLE Compounds and methods for immunotherapy and diagnosis of
 tuberculosis
 JOURNAL Patent: US 6592877-A 106 15-JUL-2003;
 FEATURES
 source Location/Qualifiers
 1. 3058
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN
 Query Match 51.1%; Score 1169; DB 6; Length 3058;
 Best Local Similarity 99.6%; Pred. No. 2,7e-135;
 Matches 1172; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 463 TCATGTTGGATTTCGGGCGCTTACCACCGGAGATCACTCCGCGAGATGATCGCGGCC 522
 Db 391 TAATGGTGGATTTCGGGCGCTTACCACCGGAGATCACTCCGCGAGATGATCGCGGCC 450
 QY 523 CGGGTTCCGCTCGCTGGTGGCGCGCTCAGATGTGGGACAGCGTGGCGAGTGAACCTGT 582
 Db 451 CGGGTTCCGCTCGCTGGTGGCGCGCTCAGATGTGGGACAGCGTGGCGAGTGAACCTGT 510
 QY 583 TTTCGGCGCGGTGGCGCTTTCAGTCGGTGTCTGGGCTCTGACGGTGGGTCGTGATGAG 642
 Db 511 TTTCGGCGCGGTGGCGCTTTCAGTCGGTGTCTGGGCTCTGACGGTGGGTCGTGATGAG 570
 QY 643 GTTCGTTCGGCGGTCTGATGTGGCGCGCTCGCGGTATGTGGCGTGGATGAGCGTCA 702
 Db 571 GTTCGTTCGGCGGTCTGATGTGGCGCGCTCGCGGTATGTGGCGTGGATGAGCGTCA 630
 QY 703 CCGCGGGGACGCGGAGCTGACCGCGCGCGCTGAGTGGCGGTGTGGCGCGCGCTACGAGA 762
 Db 631 CCGCGGGGACGCGGAGCTGACCGCGCGCGCTGAGTGGCGGTGTGGCGCGCGCTACGAGA 690
 QY 763 CGGGTATGGCTGACGCTGCGCGCGCGCTGAGTGGCGGACCGTGTGTAACCTATCA 822
 Db 691 CGGGTATGGCTGACGCTGCGCGCGCGCTGAGTGGCGGACCGTGTGTAACCTATCA 750
 QY 823 TTCTGATGAGGACCAACTTGGGGCAAAACACCGCGGATCGCGGTCAACGAGCGCG 882

Db 751 TTCTGATAGCGACCAACCTCTTGGGGCAAAACACCGCGCGATCGCGGTCAACGAGGCGG 810
 QY 883 AATACGGGAGATGTGGGCCCAAGACGCGCGCGGATGTTTGGCTACGCCCGCGGACGG 942
 Db 811 AATACGGGAGATGTGGGCCCAAGACGCGCGCGGATGTTTGGCTACGCCCGCGGACGG 870
 QY 943 CGACGGGACGCGGACGCTTGTCCGTTTCGAGGAGGCGCGGAGATGACACGCGGGTG 1002
 Db 871 CGACGGGACGCGGACGCTTGTCCGTTTCGAGGAGGCGCGGAGATGACACGCGGGTG 930
 QY 1003 GGCTCTCTCGACGAGCCCGCGGTCGAGGAGGCTCCGACACCGCGCGCGCAACCACT 1062
 Db 931 GGCTCTCTCGACGAGCCCGCGGTCGAGGAGGCTCCGACACCGCGCGCGCAACCACT 990
 QY 1063 TGATGAACAATGTGCGCCAGGCGCTGCAAGCTGGCCAGCCACCGAGGGACACGCG 1122
 Db 991 TGATGAACAATGTGCGCCAGGCGCTGCAAGCTGGCCAGCCACCGAGGGACACGCG 1050
 QY 1123 CTTCCTTCAAGCTGGTGGCTGTGGAAAGACGGTCTCGCCGATCGGTCGCGGATCAGCA 1182
 Db 1051 CTTCCTTCAAGCTGGTGGCTGTGGAAAGACGGTCTCGCCGATCGGTCGCGGATCAGCA 1110
 QY 1183 ACATGGTTCGATGGCCCAACCAACCAATGTTCGATGACCACTCGGGTGTGTCGATGACCA 1242
 Db 1111 ACATGGTTCGATGGCCCAACCAACCAATGTTCGATGACCACTCGGGTGTGTCGATGACCA 1170
 QY 1243 ACACCTTCAGCTCGATGTTGAAGGGCTTTGCTCCGGCGGCGCGCCAGGCCCTGCAAA 1302
 Db 1171 ACACCTTCAGCTCGATGTTGAAGGGCTTTGCTCCGGCGGCGCGCCAGGCCCTGCAAA 1230
 QY 1303 CGCGCGCAAAACCGGGTCCGGCGATGAGCTCGCTGGGCGAGTCTGCTGGGTTCTTCGG 1362
 Db 1231 CGCGCGCAAAACCGGGTCCGGCGATGAGCTCGCTGGGCGAGTCTGCTGGGTTCTTCGG 1290
 QY 1363 GTCTGGCGGTGGGTGGCGCGCAACTTGGTTCGGGCGGCTCGGTGGTTCGTTTCGG 1422
 Db 1291 GTCTGGCGGTGGGTGGCGCGCAACTTGGTTCGGGCGGCTCGGTGGTTCGTTTCGG 1350
 QY 1423 TGCCGACGCTGGGCGCGCGCCCAACAGGAGTCACTCCCGCGCGCGCGCGCTGCGCG 1482
 Db 1351 TGCCGACGCTGGGCGCGCGCCCAACAGGAGTCACTCCCGCGCGCGCGCGCTGCGCG 1410
 QY 1483 TGACACGCTGACACAGCCCGCGGAAAGAGGCGCGCGGATGCTGGGCGGCTGCGCG 1542
 Db 1411 TGACACGCTGACACAGCCCGCGGAAAGAGGCGCGGATGCTGGGCGGCTGCGCG 1470
 QY 1543 TGGGCGACATGGCGCGCGGCGCGCTGTTGGCTCAGTGGTGTCTGCTGTTCCGCGC 1602
 Db 1471 TGGGCGACATGGCGCGCGGCGCGCTGTTGGCTCAGTGGTGTCTGCTGTTCCGCGC 1530
 QY 1603 GACCTATGTGATGCGGCACTTCTCCGCGAGCGCGCA 1639
 Db 1531 GACCTATGTGATGCGGCACTTCTCCGCGAGCGCGCA 1567

RESULT 15
 AX429698
 LOCUS
 DEFINITION Sequence 106 from Patent EP1203817.
 ACCESSION AX429698
 VERSION AX429698.1 GI:21540898
 KEYWORDS
 SOURCE unidentified
 ORGANISM unidentified
 unclassified.
 REFERENCE 1
 AUTHORS Read, S.G., Skeiky, Y.A., Dillon, D.C., Campos-Neto, A., Houghton, R.L.,
 Vedvick, T.S. and Twardzik, D.R.
 TITLE Compounds and methods for immunotherapy and diagnosis of
 tuberculosis
 JOURNAL Patent: EP 1203817-A 106 08-MAY-2002;
 FEATURES
 Location/Qualifiers

source	1. .3058		
	/organism="unidentified"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:32644"		
ORIGIN			
Query Match	51.1%; Score 1169; DB 6; Length 3058;		
Best Local Similarity	99.6%; Pred. No. 2.7e-135;		
Matches 1172; Conservative	0; Mismatches 5; Indels 0; Gaps 0;		
Qy	463	TCATGTCGATTTTCGGGGCGTTTACACCGAGAGATCAACTCCGCGAGGATGTACGCCGCC	522
Db	391	TAATGTCGATTTTCGGGGCGTTTACACCGAGAGATCAACTCCGCGAGGATGTACGCCGCC	450
Qy	523	CGGTTTCGGCTTCGCTGGTGGCGCGCTCAGATGTGGACACGCTGGCGAGTGACTCTGT	582
Db	451	CGGTTTCGGCTTCGCTGGTGGCGCGCTCAGATGTGGACACGCTGGCGAGTGACTCTGT	510
Qy	583	TTTCGGCGCGCTCGCGCTTTTCAGTCCGTGCTCTGGGGTCTGACGCTGGGGTCTGTGATAG	642
Db	511	TTTCGGCGCGCTCGCGCTTTTCAGTCCGTGCTCTGGGGTCTGACGCTGGGGTCTGTGATAG	570
Qy	643	GTTCGTGCGCGGCTCTGATGTGGCGCGGCTCTCGCGTATGTGGGTGATGAGCGTCA	702
Db	571	GTTCGTGCGCGGCTCTGATGTGGCGCGGCTCTCGCGTATGTGGGTGATGAGCGTCA	630
Qy	703	CCGCGGGCAGCGCGAGCTGACCGCCCGCAGGTCCCGGTTGCTGGCGCGGCTACGAGA	762
Db	631	CCGCGGGCAGCGCGAGCTGACCGCCCGCAGGTCCCGGTTGCTGGCGCGGCTACGAGA	690
Qy	763	CGCGTATGGGCTGACGGTGCCTCCCGCGGCTGATTCGCCAGAACCGTGTGAACTGATGA	822
Db	691	CGCGTATGGGCTGACGGTGCCTCCCGCGGCTGATTCGCCAGAACCGTGTGAACTGATGA	750
Qy	823	TTCTGATAGCAGCAACCTCTTGGGGCAAAACACCCCGCGATCGCGGTCAACGAGGCGG	882
Db	751	TTCTGATAGCAGCAACCTCTTGGGGCAAAACACCCCGCGATCGCGGTCAACGAGGCGG	810
Qy	883	AATACGGCGAGATGTGGGCCCAAGACCGCGCGGATGTTTGGCTACGCCGGCGGCAACG	942
Db	811	AATACGGCGAGATGTGGGCCCAAGACCGCGCGGATGTTTGGCTACGCCGGCGGCAACG	870
Qy	943	CGACGGGACGGGAGCTTGTCTCGGTTCCAGAGGCGCGGAGATGACGAGCGGGGTG	1002
Db	871	CGACGGGACGGGAGCTTGTCTCGGTTCCAGAGGCGCGGAGATGACGAGCGGGGTG	930
Qy	1003	GGCTCTCTCGAGCAGCGCCCGCGGTCGAGAGGCGCTCCGACACCGCGCGCGGCAACCACT	1062
Db	931	GGCTCTCTCGAGCAGCGCCCGCGGTCGAGAGGCGCTCCGACACCGCGCGGCGGCAACCACT	990
Qy	1063	TGATGAACAATGTGCCCGCAGCGCTGCAACAGCTGGCCCHGCCACGAGGGCACCCAGC	1122
Db	991	TGATGAACAATGTGCCCGCAGCGCTGCAACAGCTGGCCCGCCACGAGGGCACCCAGC	1050
Qy	1123	CTTCTTCCAAAGCTGGGTGGCTGTGGAAGACGCTCTCGCGCATCGGTCCCGATCAGCA	1182
Db	1051	CTTCTTCCAAAGCTGGGTGGCTGTGGAAGACGCTCTCGCGCATCGGTCCCGATCAGCA	1110
Qy	1183	ACATGGTGTGATGGCCAAACACACATGTCGATGACCAACTCGGGTGTGATGACCA	1242
Db	1111	ACATGGTGTGATGGCCAAACACACATGTCGATGACCAACTCGGGTGTGATGACCA	1170
Qy	1243	ACACCTTGAGCTTCGATGTTGAAGGCTTTTGTCTCGGGCGGCGCGCGCGCGTGCMAA	1302
Db	1171	ACACCTTGAGCTTCGATGTTGAAGGCTTTTGTCTCGGGCGGCGCGCGCGCGTGCMAA	1230
Qy	1303	CCGCGCGCAAAACCGGGTCCGGCGATGAGCTCGCTGGGCGAGCTCGCTGGGTTCTTCGG	1362
Db	1231	CCGCGCGCAAAACCGGGTCCGGCGATGAGCTCGCTGGGCGAGCTCGCTGGGTTCTTCGG	1290
Qy	1363	GTCTGGGGGTGGGGTGGCGCCCAACTTGGGTGGGGCGGCTCGGTGCGTTGTTGCGG	1422
Db	1291	GTCTGGGGGTGGGGTGGCGCCCAACTTGGGTGGGGCGGCTCGGTGCGTTGTTGCGG	1350

Qy	1423	TGCCGCAAGGCTTGGGCGCGGCCCAACACAGGCAATCACCCCGGCGCGGGGCTGCCGC	1482
Db	1351	TGCCGCAAGGCTTGGGCGCGGCCCAACACAGGCAATCACCCCGGCGCGGGGCTGCCGC	1410
Qy	1483	TGACCAAGCTTGACACAGCGCGCGGAAAGAGGGCCCGGAGATGCTGGCGCGGCTGCCCG	1542
Db	1411	TGACCAAGCTTGACACAGCGCGCGGAAAGAGGGCCCGGAGATGCTGGCGCGGCTGCCCG	1470
Qy	1543	TGGGCGAGATGGGCGCGCAGGCGCGGTGGTGGGCTCAGTGGTGTGCTGGTGTTCGCGCGC	1602
Db	1471	TGGGCGAGATGGGCGCGCAGGCGCGGTGGTGGGCTCAGTGGTGTGCTGGTGTTCGCGCGC	1530
Qy	1603	GACCTATGTGATGCCGCAATTTCCGGCAGCGCGGA	1639
Db	1531	GACCTATGTGATGCCGCAATTTCCGGCAGCGCGGCTA	1567

Search completed: July 3, 2004, 20:19:30
Job time : 8840 secs

Search completed: July 3, 2004, 20:19:30
Job time : 8840 secs

Qy	1423	TGCCGAGGCTTGGGCGCGCCCAACAGGAGTCAACCCGCGGCGCGGCGCTGCGGC	1482
Db	1351	TGCCGAGGCTTGGGCGCGCCCAACAGGAGTCAACCCGCGGCGCGGCGCTGCGGC	1410
Qy	1483	TGACCAAGCTTACCAAGCGCCCGGAAAGAGGGCCCGGCGAGATGCTGGGCGGGGTGCGCG	1542
Db	1411	TGACCAAGCTTACCAAGCGCCCGGAAAGAGGGCCCGGCGAGATGCTGGGCGGGGTGCGCG	1470
Qy	1543	TGGGGCAGATGGGCGCGCAGGGCCCGTGGTGGGCTCAGTGTGCTGGTGTTCGCGCGC	1602
Db	1471	TGGGGCAGATGGGCGCGCAGGGCCCGTGGTGGGCTCAGTGTGCTGGTGTTCGCGCGC	1530
Qy	1603	GACCTTATGTGATCCCGCATTTCTCCGGCAGCGCGCGA	1639
Db	1531	GACCTTATGTGATCCCGCATTTCTCCGGCAGCGCGCTA	1567

1. The first part of the document is a letter from the President of the United States to the Congress, dated January 1, 1861.

2. The second part is a report from the Secretary of the Treasury, dated January 1, 1861.

3. The third part is a report from the Secretary of the Interior, dated January 1, 1861.

4. The fourth part is a report from the Secretary of the Navy, dated January 1, 1861.

5. The fifth part is a report from the Secretary of the War, dated January 1, 1861.

6. The sixth part is a report from the Secretary of the State, dated January 1, 1861.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2004, 14:04:57 ; Search time 899 Seconds
(without alignments)
10807.151 Million cell updates/sec

Title: US-09-597-796c-11
Perfect score: 2287
Sequence: 1 tetagaataattttgttta.....ggntgtacaagccgaaa 2287

Scoring table: IDENTITY_NUC
Gapop 10_0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database : N_Geneseq_29Jan04:*
- 1: Geneseqn1980s:*
 - 2: Geneseqn1990s:*
 - 3: Geneseqn2000s:*
 - 4: Geneseqn2001as:*
 - 5: Geneseqn2001bs:*
 - 6: Geneseqn2002s:*
 - 7: Geneseqn2003as:*
 - 8: Geneseqn2003bs:*
 - 9: Geneseqn2003cs:*
 - 10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2284	99.9	2287	6	AAD47083 Mycobacte
2	2284	99.9	2287	6	AAD28342 Mycobacte
3	2266.4	99.1	2287	2	AAD20194 Mycobacte
4	2256.2	98.7	2286	6	Abk14128 DNA encod
5	2187.4	95.6	2191	4	AAL40773 Nucleotid
6	2186.8	95.6	2451	8	ADA26360 Mycobacte
7	2186.8	95.6	2457	8	ADA26359 Mycobacte
8	2186.8	95.6	2637	8	ADA26358 Mycobacte
9	2186.8	95.6	2808	6	AAD47110 Mycobacte
10	2186.8	95.6	3060	8	ADA26363 M. bovis
11	2186.8	95.6	3104	8	ADA26362 Mycobacte
12	2186.8	95.6	3474	8	ADA26361 Mycobacte
13	2186.8	95.6	2190	6	AAD47084 Mycobacte
14	2185.2	95.5	2190	6	AAD28343 Mycobacte
15	2185.2	95.5	1797	2	AAD20205 Mycobacte
16	1770.8	77.4	1797	6	AAD47086 Mycobacte
17	1770.8	77.4	1797	6	AAD28344 Mycobacte
18	1770.8	77.4	1797	6	Abk14139 DNA encod
19	1770.8	77.4	1797	6	Abk14139 Mycobacte
20	1582	69.2	3030	8	ADA26355 Mycobacte
21	1572.6	68.8	2181	8	ADA26353 Mycobacte
22	1183.2	51.7	2445	5	AAS03792 M. tuberc
23	1175.4	51.4	2365	5	AAS03791 M. tuberc

24	1175.2	51.4	2232	5	AAS03790 M. tuberc
25	1172.8	51.3	1801	2	AAD20203 Mycobacte
26	1169	51.1	3058	2	AAV44395 Mycobacte
27	1169	51.1	3058	2	AAV64503 M. tuberc
28	1169	51.1	3058	2	AZ19093 M. tuberc
29	1169	51.1	3058	2	AZ19093 M. tuberc
30	1169	51.1	3058	5	AAS03779 M. tuberc
31	1169	51.1	3058	6	AAD47082 Mycobacte
32	1169	51.1	3058	6	AAD28341 Mycobacte
33	1169	51.1	75216	6	ABX09141 Mycobacte
34	1169	51.1	110000	4	Continuation (14 o
35	1169	51.1	110000	4	Continuation (14 o
36	1168.6	51.1	1176	2	AAX34030 Mycobacte
37	1167.4	51.0	1188	2	AAX34031 Mycobacte
38	1164.2	50.9	1565	2	AAT91455 Mycobacte
39	1164.2	50.9	1565	2	AAT91521 Mycobacte
40	1160.8	50.8	1800	6	ABK14137 DNA encod
41	955	41.8	110000	4	Continuation (39 o
42	947.4	41.4	110000	4	Continuation (16 o
43	942.6	41.2	110000	4	Continuation (16 o
44	941	41.1	3027	2	AAV44397 Mycobacte
45	941	41.1	3027	2	AAV64506 M. tuberc

ALIGNMENTS

RESULT 1
AAD47083
ID AAD47083 standard; DNA; 2287 BP.
XX AAD47083;
AC AAD47083;
XX 29-AUG-2003 (revised)
DT 27-JAN-2003 (first entry)
XX Mycobacterium sp. MTB72F fusion protein encoding DNA.
DE Vaccine; immunity; diagnostic agent; gene therapy; TBH9; antigen; Ra35;
KW Ra12; MTB72F; chimeric; gene; ds.
XX Mycobacterium sp.
OS Mycobacterium tuberculosis.
OS Chimeric.

XX Key Location/Qualifiers
FT CDS 42..2231
FT /product= "MTB72F fusion protein"
XX WO200272792-A2.
XX 19-SEP-2002.
XX 13-MAR-2002; 2002WO-US008223.
XX 13-MAR-2001; 2001US-0275837P.
XX (CORI-) CORIYA CORP.
XX skeiky Y, Brannon M, Guderian J;
XX WPI; 2002-759844/82.
XX P-PSDB; AAE29708.

XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
tuberculosis.
XX Disclosure; Page 87-90; 155pp; English.
PS The invention relates to a recombinant nucleic acid molecule encoding a
CC

541 TGGCCGCGCTCAGATGAGGACAGCGTGGAGTGAACCTGTTTTCGGCCGCGTTCGGCGT 600
541 TGGCCGCGCTCAGATGAGGACAGCGTGGAGTGAACCTGTTTTCGGCCGCGTTCGGCGT 600
601 TTCAATCGGTGCTGGGGTCTGACGGTGGGGTCTGAGTGGTCTGAGTGGTCTGAGTGGTCTGA 660
601 TTCAATCGGTGCTGGGGTCTGACGGTGGGGTCTGAGTGGTCTGAGTGGTCTGAGTGGTCTGA 660
661 TGGTGGCGCGGCTCGCCGATGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 720
661 TGGTGGCGCGGCTCGCCGATGATGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 720
721 TGAACCGCGCGGCTCGGGTCTGCTCGCGGCTTACGAGCGGCTTACGAGCGGCTTACGAGCGG 780
721 TGAACCGCGCGGCTCGGGTCTGCTCGCGGCTTACGAGCGGCTTACGAGCGGCTTACGAGCGG 780
781 TGGCCCGCGGCTGATCGCGAGAACCGTCTGACTGATGATGATGATGATGATGATGATGATGATG 840
781 TGGCCCGCGGCTGATCGCGAGAACCGTCTGACTGATGATGATGATGATGATGATGATGATGATG 840
841 TCTTGGGGCAAAAACACCCCGCGATCGCGTCAACGAGGCGGATACGCGGAGATGTTGGG 900
841 TCTTGGGGCAAAAACACCCCGCGATCGCGTCAACGAGGCGGATACGCGGAGATGTTGGG 900
901 CCGAGAGCGCGCGATGTTGGCTACGCGCGGCGGACGCGGACGCGGACGTTGGG 960
901 CCGAGAGCGCGCGATGTTGGCTACGCGCGGCGGACGCGGACGCGGACGTTGGG 960
961 TGCTGCGGCTCGAGAGCGCGGAGATGACAGCGCGGCTGCTTCCAGAGCGCGG 1020
961 TGCTGCGGCTCGAGAGCGCGGAGATGACAGCGCGGCTGCTTCCAGAGCGCGG 1020
1021 CCGCGTCTGAGAGGCTTCCGACACCGCGCGGCGAACCAAGTTGATGAACAATGTGCC 1080
1021 CCGCGTCTGAGAGGCTTCCGACACCGCGCGGCGAACCAAGTTGATGAACAATGTGCC 1080
1081 AGGCGTCTGACAGCTGCGCGACCGCGGACGAGGCGGACGCGGCTTCCAGAGCGG 1140
1081 AGGCGTCTGACAGCTGCGCGACCGCGGACGAGGCGGACGCGGCTTCCAGAGCGG 1140
1141 GCCTGTGGAAGAGGCTCGCGGATCGGTCGCGGATCAGCAACATGTTGATGAACAATGTGCC 1200
1141 GCCTGTGGAAGAGGCTCGCGGATCGGTCGCGGATCAGCAACATGTTGATGAACAATGTGCC 1200
1201 ACACACATGTCGATGACCACTCGGCTGTCGATGATGATGATGATGATGATGATGATGATGATG 1260
1201 ACACACATGTCGATGACCACTCGGCTGTCGATGATGATGATGATGATGATGATGATGATGATG 1260
1261 TGAAGGCTTTGCTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
1261 TGAAGGCTTTGCTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
1321 TCGCGCGATGAGTCTCGTGGGAGTCTGCTGGGTTCTCGGCTCTGGGCTCTGGGCTCTGGGCT 1380
1321 TCGCGCGATGAGTCTCGTGGGAGTCTGCTGGGTTCTCGGCTCTGGGCTCTGGGCTCTGGGCT 1380
1381 CCGCCAACTTGGGTCTGGGCGGCTCGGTCGTTGTCGTTGTCGTTGTCGTTGTCGTTGTCGTTG 1440
1381 CCGCCAACTTGGGTCTGGGCGGCTCGGTCGTTGTCGTTGTCGTTGTCGTTGTCGTTGTCGTTG 1440
1441 CCGCCAACTTGGGTCTGGGCGGCTCGGTCGTTGTCGTTGTCGTTGTCGTTGTCGTTGTCGTTG 1500
1441 CCGCCAACTTGGGTCTGGGCGGCTCGGTCGTTGTCGTTGTCGTTGTCGTTGTCGTTGTCGTTG 1500
1501 CCGCGAAGAGGCGCGGCGGATGCTGGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCGGCG 1560
1501 CCGCGAAGAGGCGCGGCGGATGCTGGGCGGCTGCGGCGGCTGCGGCGGCTGCGGCGGCGGCG 1560
1561 GGGCCGCTGCTGGGCTCAGTGGTGTCTGCTGTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCG 1620
1561 GGGCCGCTGCTGGGCTCAGTGGTGTCTGCTGTTCCGCGGCGGCGGCGGCGGCGGCGGCGGCG 1620
1621 ATTCTCGGCGCGGCGGATATCGCCCGCGGCGGCTTGTGCGAGGACGCGGTTGCGCGGACT 1680

1621 ATTCTCGGCGCGGCGGATATCGCCCGCGGCGGCTTGTGCGAGGACGCGTTTCGCGGACT 1680
1681 TCCCGCGCTGCGGCTCGACCGGTCGCGGATGTCGCCAAAGTGGGGCCACAGTGGTCA 1740
1681 TCCCGCGCTGCGGCTCGACCGGTCGCGGATGTCGCCAAAGTGGGGCCACAGTGGTCA 1740
1741 ACATCAACACAACTGCGGCTACAAACAGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800
1741 ACATCAACACAACTGCGGCTACAAACAGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800
1801 ATCCCAACCGGTCGTCGTCGACCAACACCTACGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGG 1860
1801 ATCCCAACCGGTCGTCGTCGACCAACACCTACGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGG 1860
1861 GCTTCAGCGTCGCGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1920
1861 GCTTCAGCGTCGCGCTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1920
1921 AGGATGTCGCGGTCGTCGTCGACGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1980
1921 AGGATGTCGCGGTCGTCGTCGACGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1980
1981 GCGGCGTCGCGGTCGTCGTCGACGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2040
1981 GCGGCGTCGCGGTCGTCGTCGACGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2040
2041 GCGCGGTCGCGGTCGTCGTCGACGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2100
2041 GCGCGGTCGCGGTCGTCGTCGACGCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2100
2101 CGCTGACCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2160
2101 CGCTGACCGGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2160
2161 CCGGTGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2220
2161 CCGGTGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2220
2221 CCGGTGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2280
2221 CCGGTGATTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2280
2281 CCGGCGG 2287
2281 CCGGCGG 2287

RESULT 3
AAZ20194
ID AAZ20194 standard; DNA; 2287 BP.
XX
AC AAZ20194;
XX
DT 17-JAN-2000 (first entry)
XX
DE Mycobacterium tuberculosis antigen fusion protein Mtb32A DNA.
XX
KW Tuberculosis; antigen; fusion protein; Mtb32A; Ra12; TBH9; Ra35;
XX diagnosis; therapy; vaccine; immunogen; ss.
XX
OS Mycobacterium tuberculosis.
XX
PH Key Location/Qualifiers
CDS 42..2231
FT /*tag= a
XX
PN WO9951748-A2
XX
PD 14-OCT-1999.
XX
PF 07-APR-1999;
XX 99WO-US007717.

Db 1621 ATTCTCCGCGACCGCGGATATCGCCCGCGCGCTTGTGCGAGGACCGGTTCCCGCACT 1680

Qy 1681 TCCCGCGCTGCCCTCGTCCAGCGCTCGCGGATGTCGCCCAAGTGGGCGCACAGTGTCTCA 1740

Db 1681 TCCCGCGCTGCCCTCGTCCAGCGCTCGCGGATGTCGCCCAAGTGGGCGCACAGTGTCTCA 1740

Qy 1741 ACATCAACACCAAACTGGGTGTACAAACACGCGTGGGCGCGGACCGGATCGTCTATCG 1800

Db 1741 ACATCAACACCAAACTGGGTGTACAAACACGCGTGGGCGCGGACCGGATCGTCTATCG 1800

Qy 1801 ATCCCAACGCTGTGCTGTGACCAACACACGATGATCGCGGCGCGCACGACATCAATG 1860

Db 1801 ATCCCAACGCTGTGCTGTGACCAACACACGATGATCGCGGCGCGCACGACATCAATG 1860

Qy 1861 CGTTCCAGCGTCCGCTCCCGCCAAACCTACGCGCTCGATGTGCTCGGTGTATGACCGCACCC 1920

Db 1861 CGTTCCAGCGTCCGCTCCCGCCAAACCTACGCGCTCGATGTGCTCGGTGTATGACCGCACCC 1920

Qy 1921 AGGATGTCGCGTGTGCTGAGCTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTG 1980

Db 1921 AGGATGTCGCGTGTGCTGAGCTGCGCGGTGCGCGGTGCGCGGTGCGCGGTGCGCGGTG 1980

Qy 1981 CCGCGCTGCCGCTGTGAGCGCTGCTGCGGATGAGCGGTCAGTTCAGTTCGCGGTGCGCGGT 2040

Db 1981 CCGCGCTGCCGCTGTGAGCGCTGCTGCGGATGAGCGGTCAGTTCAGTTCGCGGTGCGCGGT 2040

Qy 2041 CGCCCGTGGCGTGTGCTGAGGCTGCTGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGG 2100

Db 2041 CGCCCGTGGCGTGTGCTGAGGCTGCTGCGGTCGCGGTCGCGGTCGCGGTCGCGGTCGCGG 2100

Qy 2101 CGCTGACCGGTGCGGATGAGACATGAGCGGTCAGTTCAGTTCGCGGTGCGCGGTGCGCGGT 2160

Db 2101 CGCTGACCGGTGCGGATGAGACATGAGCGGTCAGTTCAGTTCGCGGTGCGCGGTGCGCGGT 2160

Qy 2161 CCGGTGATTCGGCGGCGCGCTGCTCAACGCGCTAGGACAGGTCGCGGTGATGAACACCGG 2220

Db 2161 CCGGTGATTCGGCGGCGCGCTGCTCAACGCGCTAGGACAGGTCGCGGTGATGAACACCGG 2220

Qy 2221 CCGGTGCTAGGATATCATCACTGCGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCT 2280

Db 2221 CCGGTGCTAGGATATCATCACTGCGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCT 2280

Qy 2281 CCCGAAA 2287

Db 2281 CCCGAAA 2287

RESULT 4

ABK14128

XX ABK14128 standard; DNA; 2286 BP.

XX AC ABK14128;

XX XX

DT 29-AUG-2003 (revised)

DT 08-MAY-2002 (first entry)

XX XX

DE DNA encoding antigenic fusion protein Ra12-TbH9-Ra35 (Mtb32-Mtb39).

XX Fusion protein; tuberculosis; Mycobacterium tuberculosis; gene; ds;

KW tuberculostatic; immunogen; vaccine; Mtb32-Mtb39; Ra12; TbH9; Ra35;

KW Ra12-TbH9-Ra35.

XX XX

OS Mycobacterium tuberculosis.

XX Chimeric.

XX XX

FH Key Location/Qualifiers

FT CDS 42..2231

FT /*tag= a

FT /product= "Mtb32-Mtb39"

FT /transl_except= (pos:498..506, aa:Asn-Ala)

FT /transl_except= (pos:597..605, aa:Ala-Gln)

FT /transl_except= (pos:798..802, aa:Ala)

FT /note= "This codon has an apparent 2 nucleotide insertion"

FT which alters the reading frame"

XX US2002009459-A1.

PN 24-JAN-2002.

XX XX

PF 07-APR-1999; 99US-00287849.

XX XX

XX 13-MAR-1997; 97US-00818112.

PR 01-OCT-1997; 97US-00942578.

PR 18-FEB-1998; 98US-00025197.

PR 07-APR-1998; 98US-00056556.

PR 30-DEC-1998; 98US-00223040.

XX XX

(REED/) REED S G.

PA (SKSI/) SKEIKY Y A.

PA (DILL/) DILLON D C.

PA (ALDE/) ALDERSON M.

PA (CAMP/) CAMPOS-NETO A.

XX XX

PI Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;

XX XX

WPI: 2002-171134/22.

DR P-PSDB; AAU74588.

XX XX

PT New fusion proteins of Mycobacterium tuberculosis antigens, useful for
 PT diagnosing, treating or preventing M. tuberculosis infection,
 PT particularly as vaccine for treating or preventing tuberculosis.

XX XX

PS Example; Fig 1; 62pp; English.

XX XX

CC The invention relates to a purified polypeptide which induces an immune
 CC response of Mycobacterium tuberculosis. Polypeptides of the invention are
 CC useful for diagnosing, treating or preventing M. tuberculosis infection,
 CC particularly tuberculosis infection. In particular, the polypeptides are
 CC useful as a vaccine formulation with an adjuvant to afford long-term
 CC protection in animals against the development of tuberculosis. The
 CC protein coding sequence may be used to encode a protein product for use
 CC as an immunogen to induce and/or enhance an immune response to M.
 CC tuberculosis. This sequence represents DNA encoding an M. tuberculosis
 CC fusion protein of the invention. (Updated on 29-AUG-2003 to standardise
 CC OS field)

XX XX

SQ Sequence 2286 BP; 373 A; 719 C; 807 G; 381 T; 0 U; 6 Other;

Query Match 98.7%; Score 2256.2; DB 6; Length 2286;

Best Local Similarity 99.5%; Pred. No. 0;

Matches 2275; Conservative 0; Mismatches 11; Indels 1; Gaps 1;

Qy 1 TCTAGAAATAATTTTGTGTTTACTTTTAAAGANANATATACATGATCATCACCATC 60

Db 1 TCTAGAAATAATTTTGTGTTTACTTTTAAAGANANATATACATGATCATCACCATC 60

Qy 61 ACACGGCGGTCGCGATTAACCTTCCAGCTGTCCAGGTCGCGAGGATTCGCCATTCGGA 120

Db 61 ACACGGCGGTCGCGATTAACCTTCCAGCTGTCCAGGTCGCGAGGATTCGCCATTCGGA 120

Qy 121 TCGGGCAGGCGATGCGGATCGGGCGGCGAGATCCGATCGGGTGGGGTCAACCGGTC 180

Db 121 TCGGGCAGGCGATGCGGATCGGGCGGCGAGATCCGATCGGGTGGGGTCAACCGGTC 180

Qy 181 ATATCGGCGCTTACCGCTTCTCGGCTTGGGTGTGTCGACAAACACGCAACGGGCGAC 240

Db 181 ATATCGGCGCTTACCGCTTCTCGGCTTGGGTGTGTCGACAAACACGCAACGGGCGAC 240

Qy 241 GAGTCCAAACGGTGTGTCGGAGCGCTCCGGCGGCGAGTCTCGGATCTCCACGGGCGAC 300

Db 241 GAGTCCAAACGGTGTGTCGGAGCGCTCCGGCGGCGAGTCTCGGATCTCCACGGGCGAC 300

Qy 301 TGATCAACGGGTCGACGGCGCTCCGATCAACTCGGCCACCGGATGCGGAGCGGCTTA 360

Db 301 TGATCAACGGGTCGACGGCGCTCCGATCAACTCGGCCACCGGATGCGGAGCGGCTTA 360

QY 361 ACGGGATCATCCCGGTGACGTCATCTCGTGTGACCTGGGAAACCAAGTCGGCGGCACGC 420
Db 361 ACGGGATCATCCCGGTGACGTCATCTCGTGTGACCTGGGAAACCAAGTCGGCGGCACGC 420
QY 421 GTACAGGGAACCTGACATTGGCCGAGGACACCCCGGCCGAATTTCATGTGTGATTTCGGGG 480
Db 421 GTACAGGGAACCTGACATTGGCCGAGGACACCCCGGCCGAATTTCATGTGTGATTTCGGGG 480
QY 481 CATTACACCGAGATCAATCCCGGAGATTAACCCCGGCCCGGGTTCCGCTCGCTGG 540
Db 481 CATTACACCGAGATCAACNCCGAGATGTACCGCGGCCCGGGTTCCGCTCGCTGG 540
QY 541 TGGCCGCGCTCAGATGTGGGACAGCTGGCGAGTGACCTGTTTTCGCGCGCGTGGCGGT 600
Db 541 TGGCCGCGCTCAGATGTGGGACAGCTGGCGAGTGACCTGTTTTCGCGCGCGTGGCGGT 600
QY 601 TTCACTCGGTGTCTGGGGTCTGACGGTGGGTTCGTGTAGTGTTCGTGGCGGGTCTGA 660
Db 601 TTCACTCGGTGTCTGGGGTCTGACGGTGGGTTCGTGTAGTGTTCGTGGCGGGTCTGA 660
QY 661 TGGTGGCGGCGCTCGCTGATGTGGCTGTGATGAGCTCACCGGGGCGAGCGGAGC 720
Db 661 TGGTGGCGGCGCTCGCTGATGTGGCTGTGATGAGCTCACCGGGGCGAGCGGAGC 720
QY 721 TGAACCGCCCGCAGGTCGGGTTCGTGGCGGCGCTTACGAGCGGCTATGGGCTGAGCGG 780
Db 721 TGAACCGCCCGCAGGTCGGGTTCGTGGCGGCGCTTACGAGCGGCTATGGGCTGAGCGG 780
QY 781 TGGCCCGCGCGGTGATGCGCGAAGACCGGTGTGAATGATGATTTCTGATAGCGACCAAC 840
Db 781 TGGCCCGCGCGGTGATGCGCGG- GAACCGGTGTGAATGATGATTTCTGATAGCGACCAAC 839
QY 841 TCTTGGGCAACACCCCGGATCGCGTCAACGAGCGGAAATACGCGGAGATGTGG 900
Db 840 TCTTGGGCAACACCCCGGATCGCGTCAACGAGCGGAAATACGCGGAGATGTGG 899
QY 901 CCCAAGACGCGCGCGATTTGGCTACGCGCGCGGACGCGACGCGGACGCGGACGT 960
Db 900 CCCAAGACGCGCGCGATTTGGCTACGCGCGCGGACGCGACGCGGACGCGGACGT 959
QY 961 TGGTGGCTGAGGAGGCGCGGATGACGAGCGGCGGTGGCTTCGCGAGCGGCG 1020
Db 960 TGGTGGCTGAGGAGGCGCGGATGACGAGCGGCGGTGGCTTCGCGAGCGGCG 1019
QY 1021 CCGCGTGCAGGAGGCTCCGACACCGCGCGGCGAACAGTTGATGAACAATGTGCCCC 1080
Db 1020 CCGCGTGCAGGAGGCTCCGACACCGCGCGGCGAACAGTTGATGAACAATGTGCCCC 1079
QY 1081 AGCGGTGCAACAGCTGGCGGCGGACCGACGAGGCGACGAGCTTTTCCAGCTGGGTG 1140
Db 1080 AGCGGTGCAACAGCTGGCGGCGGACCGACGAGGCGACGAGCTTTTCCAGCTGGGTG 1139
QY 1141 GCCTGTGGAAGAGCGGTCTCGCGCATCGGTGCGCGGATGACCAACATGGTGTGAGGCCA 1200
Db 1140 GCCTGTGGAAGAGCGGTCTCGCGCATCGGTGCGCGGATGACCAACATGGTGTGAGGCCA 1199
QY 1201 ACAACACATGTGATGACCAACTCGGTTGTGATGACCAACACTTTAGCTGATGT 1260
Db 1200 ACAACACATGTGATGACCAACTCGGTTGTGATGACCAACACTTTAGCTGATGT 1259
QY 1261 TGAAGGCTTTGCTCGGCGGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
Db 1260 TGAAGGCTTTGCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1319
QY 1321 TCGGGCGATGAGCTCGCTGGGAGCTCGCTGGGTCTTCGCGGTCTGGGCGTGGGGTGG 1380
Db 1320 TCGGGCGATGAGCTCGCTGGGAGCTCGCTGGGTCTTCGCGGTCTGGGCGTGGGGTGG 1379
QY 1381 CCGCAACTTTGGGTGGGCGGCGCTCGGTGCGGTTCGTGTGCGGCGGCGGCGGCGG 1440
Db 1380 CCGCAACTTTGGGTGGGCGGCGCTCGGTGCGGTTCGTGTGCGGCGGCGGCGGCGG 1439
QY 1441 CGGCCAACGAGGAGTCAACCCCGGCGGCGGCGGCGGCTGCGGCTGACGAGCTGACAGCG 1500

RESULT 5

AAL40773

ID AAL40773 standard; DNA; 2191 BP.

XX AAL40773;

XX AC

DT 03-OCT-2002 (first entry)

XX

DE Nucleotide sequence encoding Ral2-H9-32A fusion protein.

XX

Ral2; serine protease antigen; MTB32A; Mycobacterium tuberculosis;
KW vaccine; eukaryotic; prokaryotic origin; serum antibody; fusion protein;
KW immunogen; cytokine; gene; ds.

Db 1440 CGGCCAACGAGGAGTCAACCCCGGCGGCGGCGTGCCTGACGAGCTGACAGCG 1499
QY 1501 CCGCGGAAAGAGGCGCGGCGAGATGCTGGCGGGCTGCGGGTGGGCGAGATGGGCGCCA 1560
Db 1500 CCGCGGAAAGAGGCGCGGCGAGATGCTGGCGGGCTGCGGGTGGGCGAGATGGGCGCCA 1559
QY 1561 GGGCGGTGTGGGCTCAGTGGTGTGCTGCTGGCGCGACCTATGTGTATGTCGCG 1620
Db 1560 GGGCGGTGTGGGCTCAGTGGTGTGCTGCTGGCGCGACCTATGTGTATGTCGCG 1619
QY 1621 ATTTCCCGGAGCGCGGATATCGCCCGCGGCTTGTGCGAGGACCGGTTCCCGACT 1680
Db 1620 ATTTCCCGGAGCGCGGATATCGCCCGCGGCTTGTGCGAGGACCGGTTCCCGACT 1679
QY 1681 TCCCGCGCTGCGCTCGACCGCTCGCGATGTCGCCCAAGTGGGCGCACAGTGTCTCA 1740
Db 1680 TCCCGCGCTGCGCTCGACCGCTCGCGATGTCGCCCAAGTGGGCGCACAGTGTCTCA 1739
QY 1741 ACATCAACCAAACTGGGCTACAAACCGCTGGCGCGCGGACCGGATCGTCAATCG 1800
Db 1740 ACATCAACCAAACTGGGCTACAAACCGCTGGCGCGCGGACCGGATCGTCAATCG 1799
QY 1801 ATCCAAACGCTGCTGCTGACCAACCAACAGCTGATCGGGCGCGCACCGACATCAATG 1860
Db 1800 ATCCAAACGCTGCTGCTGACCAACCAACAGCTGATCGGGCGCGCACCGACATCAATG 1859
QY 1861 GGTTCAGCTCGGCTCGCGCCAAACCTACGCGCTCGATGTGGTATGATGACCGCACCC 1920
Db 1860 GGTTCAGCTCGGCTCGCGCCAAACCTACGCGCTCGATGTGGTATGATGACCGCACCC 1919
QY 1921 AGGATGCGGCTGCTGACGCTCGCGTGGCTGCGCTGCGGCGCGATCGGCTG 1980
Db 1920 AGGATGCGGCTGCTGACGCTCGCGTGGCTGCGCTGCGGCGCGATCGGCTG 1979
QY 1981 GCGCGCTCGGCTGCTGAGCGCTGCTGCGGATGGGCAACAGCGTGGGCGGCGCGGAA 2040
Db 1980 GCGCGCTCGGCTGCTGAGCGCTTCTGCGGATGGGCAACAGCGTGGGCGGCGCGGAA 2039
QY 2041 CGCCCGTGGCGTGGCGAGGCTGGTGGCGCTCGGCGCGGCGCGGCGGCGGAT 2100
Db 2040 CGCCCGTGGCGTGGCGAGGCTGGTGGCGCTCGGCGCGGCGGCGGCGGAT 2099
QY 2101 CGCTGACCGGTGGCGAGACATTTGAACGCGTTGATCCAGTTGATGCGCGGATCCAGC 2160
Db 2100 CGCTGACCGGTGGCGAGACATTTGAACGCGTTGATCCAGTTGATGCGCGGATCCAGC 2159
QY 2161 CGGTTGATTGGCGGCGGCGCTCGTCAACGCGCTAGGACAGTGGTGGTATGAACACGG 2220
Db 2160 CGGTTGATTGGCGGCGGCGCTCGTCAACGCGCTAGGACAGTGGTGGTATGAACACGG 2219
QY 2221 CCGGCTCTAGGATATCCATCACACTGGCGCGCTCGAGCAGATCCGGTGTGAACAAAG 2280
Db 2220 CCGGCTCTAGGATATCCATCACACTGGCGCGGCGGCGGCGGCGGCGGCGGCGG 2279
QY 2281 CCGGAA 2287
Db 2280 CCGGAA 2286

QY 519 GGCCCGGGTTCCGCTCGCTGCTGGCGCGCTCAGATGTGGGACAGCGTGGCGAGTGAC 578
 Db 481 GGCCCGGGTTCCGCTCGCTGCTGGCGCGCTCAGATGTGGGACAGCGTGGCGAGTGAC 540
 QY 579 CTGTTTTCGGCGCGCTCGCTGCTGGCGCGCTCAGTGGGTCTGCGGTCTGCGGTGGGTCTG 638
 Db 541 CTGTTTTCGGCGCGCTCGCTGCTGGCGCGCTCAGTGGGTCTGCGGTCTGCGGTGGGTCTG 600
 QY 639 ATAGTTCTGCTGGCGCGCTCGCTGCTGGCGCGCTCAGTGGGTCTGCGGTGGGTCTGCGGT 698
 Db 601 ATAGTTCTGCTGGCGCGCTCGCTGCTGGCGCGCTCAGTGGGTCTGCGGTGGGTCTGCGGT 660
 QY 699 GTACCGGGGCGAGCGCGCTGACCGCGCGCGCGCGCGCTGCTGGCGCGCGCTGAC 758
 Db 661 GTACCGGGGCGAGCGCGCTGACCGCGCGCGCGCGCGCTGCTGGCGCGCGCTGAC 720
 QY 759 GAGACGGGTATGGGCTGACCGTGGCGCGCGCGCGCGCTGCTGGCGCGCGCTGAC 818
 Db 721 GAGACGGGTATGGGCTGACCGTGGCGCGCGCGCGCGCTGCTGGCGCGCGCTGAC 780
 QY 819 ATGATTCTGATAGCGACCACTCTTGGGCGAAACACCGCGCGCGCTGCTGGCGCGCGCT 878
 Db 781 ATGATTCTGATAGCGACCACTCTTGGGCGAAACACCGCGCGCGCTGCTGGCGCGCGCT 840
 QY 879 GCCGAATACGGCGAGATGTGGCGCGCGCGCGCGCGCTGCTGGCGCGCGCTGAC 938
 Db 841 GCCGAATACGGCGAGATGTGGCGCGCGCGCGCGCGCTGCTGGCGCGCGCTGAC 900
 QY 939 ACGGCG 998
 Db 901 ACGGCG 960
 QY 999 GGTGGGCTCTCGAGCG 1058
 Db 961 GGTGGGCTCTCGAGCG 1020
 QY 1059 CAGTTGATGACATGTGCG 1118
 Db 1021 CAGTTGATGACATGTGCG 1080
 QY 1119 ACGGCTTCTTCCAAAGCTGGGTGGCGCTGGAAGACGCTCTCGCGCGCGCGCGCGCGCG 1178
 Db 1081 ACGGCTTCTTCCAAAGCTGGGTGGCGCTGGAAGACGCTCTCGCGCGCGCGCGCGCGCG 1140
 QY 1179 AGCAACATGTGCGATGGCG 1238
 Db 1141 AGCAACATGTGCGATGGCG 1200
 QY 1239 ACCAACACCTTGAAGCTGGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1298
 Db 1201 ACCAACACCTTGAAGCTGGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1260
 QY 1299 CAACCG 1358
 Db 1261 CAACCG 1320
 QY 1359 TCGGCTCTGGCG 1418
 Db 1321 TCGGCTCTGGCG 1380
 QY 1419 TCGGCTCTGGCG 1478
 Db 1381 TCGGCTCTGGCG 1440
 QY 1479 CCCTGACACGCTGACCG 1538
 Db 1441 CCCTGACACGCTGACCG 1500
 QY 1539 CCGGTGGGCGAGATGGCG 1598
 Db 1501 CCGGTGGGCGAGATGGCG 1560
 QY 1599 CCGGACCCCTATGTATCG 1658

Db 1561 CCAGCGCCCTATGTATGCGCGATTCCTCGGCGAGCGCGCGATATCGCCCGCGCGCGCTTG 1620
 QY 1659 TCAGCGACCGGTTCCCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 1718
 Db 1621 TCAGCGACCGGTTCCCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 1680
 QY 1719 CAAGTGGGCGCACAGGTGCTCAACATCAACACAACTGGGCTCAACAAACCGCGCTGCGCGCTG 1778
 Db 1681 CAAGTGGGCGCACAGGTGCTCAACATCAACACAACTGGGCTCAACAAACCGCGCTGCGCGCTG 1740
 QY 1779 GCGGCGACCGGCTGCTGATCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 1838
 Db 1741 GCGGCGACCGGCTGCTGATCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 1800
 QY 1839 GCGGCGCGCACCGCATCAATCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 1898
 Db 1801 GCGGCGCGCACCGCATCAATCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTGCGCGCTG 1860
 QY 1899 GTGGTGGGCTGACCG 1958
 Db 1861 GTGGTGGGCTGACCG 1920
 QY 1959 CTGGCTGCG 2018
 Db 1921 CTGGCTGCG 1980
 QY 2019 AACAGCGGTGGCG 2078
 Db 1981 AACAGCGGTGGCG 2040
 QY 2079 CAACCGGTGCG 2138
 Db 2041 CAACCGGTGCG 2100
 QY 2139 CAGTTGCGATGCG 2198
 Db 2101 CAGTTGCGATGCG 2160
 QY 2199 CAGGTGGTGGGTGATGAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2228
 Db 2161 CAGGTGGTGGGTGATGAACACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2190

RESULT 7
 ADA26359
 ID ADA26359 standard; DNA; 2487 BP.
 XX
 AC ADA26359;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Mycobacterium MTB72F-MTI (fusion MTB83F) protein encoding DNA.
 XX
 KW ds; gene; fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 KW tuberculosis; tuberculostatic; gene therapy; vaccine.
 OS Chimeric.
 OS Mycobacterium sp.
 XX
 PH Key Location/Qualifiers
 FT CDS 4..2481
 FT /+tag= a
 FT /product= "MTB72F-MTI (fusion MTB83F) protein"
 XX
 PN WO2003070187-A2.
 XX
 PD 28-AUG-2003.
 XX
 PF 18-FEB-2003; 2003WO-US004903.
 XX
 PR 15-FEB-2002; 2002US-0357351P.
 XX

DB 1741 GCCGGGACCGGCATGTCATGATCCCAAGGTCGTGCTGCTGACCAACACCGATGATC 1800
QY 1839 GCGGGGACCGACATCAATGCGTTAGCGTCGGCTCCGGCCAAACCTACGGGTCGAT 1898
DB 1801 GCGGGGACCGACATCAATGCGTTAGCGTCGGCTCCGGCCAAACCTACGGGTCGAT 1860
QY 1899 GTGGTCGGGTATGACCGACACAGGATGTCGCGGTGCTGACGTCGCGGTGCGGTCGCG 1958
DB 1861 GTGGTCGGGTATGACCGACACAGGATGTCGCGGTGCTGACGTCGCGGTGCGGTCGCG 1920
QY 1959 CTGGCTGCGGCGGATCGGTGGGCGGTCGCGGTGCTGACGTCGCGGTGCTGACGTCGCG 2018
DB 1921 CTGGCTGCGGCGGATCGGTGGGCGGTCGCGGTGCTGACGTCGCGGTGCTGACGTCGCG 1980
QY 2019 AACAGCGGTGGCGGCGGACCGCCGTCGCGGTGCTGACGTCGCGGTGCTGACGTCGCG 2078
DB 1981 AACAGCGGTGGCGGCGGACCGCCGTCGCGGTGCTGACGTCGCGGTGCTGACGTCGCG 2040
QY 2079 CAAACCGGTGCGGCGGTCGATTCGTCGCGGTGCTGACGTCGCGGTGCTGACGTCGCG 2138
DB 2041 CAAACCGGTGCGGCGGTCGATTCGTCGCGGTGCTGACGTCGCGGTGCTGACGTCGCG 2100
QY 2139 CAGTTCGATCGCGGATCCAGCGGTCGATTCGCGGTGCTGACGTCGCGGTGCTGACGTCG 2198
DB 2101 CAGTTCGATCGCGGATCCAGCGGTCGATTCGCGGTGCTGACGTCGCGGTGCTGACGTCG 2160
QY 2199 CAGGTGTCGATGATCAACACGCGCGCGTC 2228
DB 2161 CAGGTGTCGATGATCAACACGCGCGCGTC 2190

RESULT 8

ADA26358
ID ADA26358 standard; DNA; 2637 BP.
XX
AC ADA26358;
XX
DT 20-NOV-2003 (first entry)
XX
DE Mycobacterium MTB72F-Erd14 (fusion MTB89F) protein encoding DNA.
XX
KW ds; gene; fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW tuberculosis; tuberculostatic; gene therapy; vaccine.
XX
OS Chimeric.
OS Mycobacterium sp.
XX
FH Key Location/Qualifiers
CDS 4..2631
FT /*tag= a
FT /product= "MTB72F-Erd14 (fusion MTB89F)"
XX
PN WO2003070187-A2.
XX
XX 28-AUG-2003.
XX
XX 18-FEB-2003; 2003WO-US004903.
XX
XX 15-FEB-2002; 2002US-0357351P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Gunderian J, Reed S;
XX
XX WPI: 2003-697554/66.
XX P-PSDB; ADA26358.
XX
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT and MTB39 antigens, with or without the MTB85A antigen, from a
PT Mycobacterium species, useful for diagnosing, preventing and/or treating
PT tuberculosis infection.
XX

PS Claim 84; Fig 6; 112pp; English.

XX The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The
CC present sequence is used in the exemplification of the invention.

XX SQ Sequence 2637 BP; 428 A; 840 C; 928 G; 441 T; 0 U; 0 Other;

Query Match 95.6%; Score 2186.8; DB 8; Length 2637;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 39 CATATGCAATCACCATCACATCACACGCGCGCTCCGATAACTTCCAGCTGTCCAGGGT 98

DB 1 CATATGCAATCACCATCACATCACACGCGCGCTCCGATAACTTCCAGCTGTCCAGGGT 60

QY 99 GGGCAGGATTCGCCATTCGATCGGCGAGCGATGCGGCGGCGGCGGCGGCGGCGGCGG 158

DB 61 GGGCAGGATTCGCCATTCGATCGGCGAGCGATGCGGCGGCGGCGGCGGCGGCGGCGG 120

QY 159 GGTGGGGGTTCACCCACCGCTTCATATCGGGGCTTACCGCTTCTCGGCTTGGGTGTTGTC 218

DB 121 GGTGGGGGTTCACCCACCGCTTCATATCGGGGCTTACCGCTTCTCGGCTTGGGTGTTGTC 180

QY 219 GACACACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 278

DB 181 GACACACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 240

QY 279 CTCGGCATCTCCACCGGCGGCTGATCACCGCGGCTGACCGGCGGCTGACCGGCGGCGG 338

DB 241 CTCGGCATCTCCACCGGCGGCTGATCACCGCGGCTGACCGGCGGCTGACCGGCGGCGG 300

QY 339 ACCCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 398

DB 301 ACCCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360

QY 399 CAAACCAAGTCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 458

DB 361 CAAACCAAGTCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420

QY 459 GAATTCATGTTGGATTTGGGGCGTTTACACCGGAGATCAATCCCGGAGGATGACGCC 518

DB 421 GAATTCATGTTGGATTTGGGGCGTTTACACCGGAGATCAATCCCGGAGGATGACGCC 480

QY 519 GCGCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCG 578

DB 481 GCGCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCG 540

QY 579 CTGTTTTCGCGCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCG 638

DB 541 CTGTTTTCGCGCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCGGCTCG 600

QY 639 ATAGTTTCGTCGGCGGCTGATGTCGGCGGCGGCTGATGTCGGCGGCGGCTGATGTCGG 698

DB 601 ATAGTTTCGTCGGCGGCTGATGTCGGCGGCGGCTGATGTCGGCGGCGGCTGATGTCGG 660

QY 699 GTACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 758

DB 661 GTACCCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720

QY 759 GAGAGCGGCTATGGGCTGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 818

DB 721 GAGAGCGGCTATGGGCTGACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 780

QY 819 ATGATTCTGATAGCGACCAACCTCTTGGGGCAAAACACCCCGGCGGATCGGGGTCAACG 878

DB 781 ATGATTCTGATAGCGACCAACCTCTTGGGGCAAAACACCCCGGCGGATCGGGGTCAACG 840

QY 879 GCGAATACGGGAGATGTGGGCGCAAGACGGCGCGCGATGTTTGGCTAGCGCGCGG 938
 Db 841 GCGAATACGGGAGATGTGGGCGCAAGACGGCGCGCGATGTTTGGCTAGCGCGCGG 900
 QY 939 ACGGGACGGGACGGGACGGGACGGGACGGGACGGGACGGGACGGGACGGGACGGG 998
 Db 901 ACGGGACGGGACGGGACGGGACGGGACGGGACGGGACGGGACGGGACGGGACGGG 960
 QY 999 GGTGGGCTCTTCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1058
 Db 961 GGTGGGCTCTTCGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020
 QY 1059 CAGTTGATGAACAATGTCGCCAGCGGCTGCAACAGCTGGCCAGCCCAACAGCGGAC 1118
 Db 1021 CAGTTGATGAACAATGTCGCCAGCGGCTGCAACAGCTGGCCAGCCCAACAGCGGAC 1080
 QY 1119 AGCGCTTCTTCAAGCTGGGTCGCTGTGGAGACGGTCTCGCCGATCGTTCGCGATC 1178
 Db 1081 AGCGCTTCTTCAAGCTGGGTCGCTGTGGAGACGGTCTCGCCGATCGTTCGCGATC 1140
 QY 1179 AGCAATGCTGTGATGCGCAACCAACACATGTCGATGACCACTCGGCTGTGCGATG 1238
 Db 1141 AGCAATGCTGTGATGCGCAACCAACACATGTCGATGACCACTCGGCTGTGCGATG 1200
 QY 1239 ACCAATGCTGTGATGCGCAACCAACACATGTCGATGACCACTCGGCTGTGCGATG 1298
 Db 1201 ACCAATGCTGTGATGCGCAACCAACACATGTCGATGACCACTCGGCTGTGCGATG 1260
 QY 1299 CAAAACGGCGGCAAAACGGGTCGCGGATGAGCTGCTGGGAGCTGCTGGGAGCTGCT 1358
 Db 1261 CAAAACGGCGGCAAAACGGGTCGCGGATGAGCTGCTGGGAGCTGCTGGGAGCTGCT 1320
 QY 1359 TCGGGTCTGGGCGGTGGGTCGCGGCACTTGGGTCGGGCGGCTCGGTCGGTTCGTTG 1418
 Db 1321 TCGGGTCTGGGCGGTGGGTCGCGGCACTTGGGTCGGGCGGCTCGGTCGGTTCGTTG 1380
 QY 1419 TCGGTGCGGAGCTGGGTCGCGGCACTTGGGTCGGGCGGCTCGGTCGGTTCGTTG 1478
 Db 1381 TCGGTGCGGAGCTGGGTCGCGGCACTTGGGTCGGGCGGCTCGGTCGGTTCGTTG 1440
 QY 1479 CGCTGACAGCTGACGAGCGCGGCAAAAGGGCCCGGCGAGATGCTGGGCGGCTG 1538
 Db 1441 CGCTGACAGCTGACGAGCGCGGCAAAAGGGCCCGGCGAGATGCTGGGCGGCTG 1500
 QY 1539 CGGTGCGGAGATGGGCGCGGCAAAAGGGCCCGGCGAGATGCTGGGCGGCTG 1598
 Db 1501 CGGTGCGGAGATGGGCGCGGCAAAAGGGCCCGGCGAGATGCTGGGCGGCTG 1560
 QY 1599 CGCGACCTATGTGATGCGCATTTCTCGGAGCGCGGATATCGCCCGCGCGGCTG 1658
 Db 1561 CGCGACCTATGTGATGCGCATTTCTCGGAGCGCGGATATCGCCCGCGCGGCTG 1620
 QY 1659 TCGAGGACCGGTTCCGCGACTTCCCGGCTGCGGCTCGGCGGCTCGGCGGCTG 1718
 Db 1621 TCGAGGACCGGTTCCGCGACTTCCCGGCTGCGGCTCGGCGGCTCGGCGGCTG 1680
 QY 1719 CAAATGGGCGCACAGGTGGTCAACATCAACCAAACTGGGCTACAAACGCGGCGG 1778
 Db 1681 CAAATGGGCGCACAGGTGGTCAACATCAACCAAACTGGGCTACAAACGCGGCGG 1740
 QY 1779 GCGGAGCGGATGCTGATCGATCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1838
 Db 1741 GCGGAGCGGATGCTGATCGATCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
 QY 1839 GCGGAGCGGATGCTGATCGATCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1898
 Db 1801 GCGGAGCGGATGCTGATCGATCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1860
 QY 1899 GTGGTCGGGTATGACCGCAACCGAGATGTGCGGTCGCTGCTGCTGCTGCTGCTGCTGCT 1958
 Db 1861 GTGGTCGGGTATGACCGCAACCGAGATGTGCGGTCGCTGCTGCTGCTGCTGCTGCTGCT 1920
 QY 1959 CTGCGCTCGGCGGATCGGTGGGCGGCTCGCGTGGTGGTGGTGGTGGTGGTGGTGGG 2018

Db 1921 CTGCGCTCGGCGGATCGGTGGGCGGCTCGGCTGGTGGTGGTGGTGGTGGTGGG 1980
 QY 2019 AACAGCGGTGGGAGAGCGGCGGAGCGGCGGCTGCGGTCGCTGCGGAGGTGGTGGGCTCGG 2078
 Db 1981 AACAGCGGTGGGAGAGCGGCGGAGCGGCGGCTGCGGTCGCTGCGGAGGTGGTGGGCTCGG 2040
 QY 2079 CAAACGCTGCGGAGCGGCTGCGGATTCGCTGACCGGTGCGGAGAGAGACATTGAACGGGTTGATC 2138
 Db 2041 CAAACGCTGCGGAGCGGCTGCGGATTCGCTGACCGGTGCGGAGAGAGACATTGAACGGGTTGATC 2100
 QY 2139 CAGTTGATGCGGAGATCCAGCGGCTGCGGATTCGCTGAGCGGCGGCGGCGGCTGCTCAACGGCTAGGA 2198
 Db 2101 CAGTTGATGCGGAGATCCAGCGGCTGCGGATTCGCTGAGCGGCGGCGGCGGCTGCTCAACGGCTAGGA 2160
 QY 2199 CAGGTGCGGATGAACACGCGCGGCTCC 2228
 Db 2161 CAGGTGCGGATGAACACGCGCGGCTCC 2190

RESULT 9

AAD47110
 ID AAD47110 standard; DNA; 2808 BP.

XX AAD47110; AC

XX 29-AUG-2003 (revised)

DT 27-JAN-2003 (first entry)

XX Mycobacterium sp. MTB72F-Leishmania sp. MAPS (aka r95f)-fusion DNA.

XX Vaccine; immunity; diagnostic agent; gene therapy; MTB72F; MAPS;

XX chimeric; gene; ds.

OS Mycobacterium sp.

OS Leishmania sp.

OS Chimeric.

PH Key Location/Qualifiers

FT CDS 4..2796

FT /*tag= a

FT /product= "Mycobacterium sp. MTB72F-Leishmania sp. MAPS

FT (aka r95f) fusion protein"

XX WO200272792-A2.

XX 19-SEP-2002.

XX 13-MAR-2002; 2002WO-US008223.

XX 13-MAR-2001; 2001US-0275837P.

XX (CORI-) CORIXA CORP.

PA Skeiky Y, Brannon M, Guderian J;

PI WPI; 2002-759844/82.

XX P-PSDB; AAE29731.

XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeifP, M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity against pathogenic microorganisms e.g. Leishmania and Mycobacterium tuberculosis.

PS Example 6; Page 128-129; 155pp; English.

XX The invention relates to a recombinant nucleic acid molecule encoding a fusion polypeptide. The recombinant nucleic acid comprises a heterologous polynucleotide sequence encoding an antigen or an antigenic fragment from Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a polypeptide or its fragment. The Leishmania polynucleotide is selected from TSA, LeifP, M15, and 6H polynucleotides. Sequences of the invention are used in methods for eliciting immune response in mammals. They are

CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is
 CC Mycobacterium sp. MTB72F-Leishmania sp. thiol specific antioxidant [TSA;
 CC MAPS (aka r95f)] fusion DNA. This sequence comprises Mycobacterium sp.
 CC MTB72F (a 72 kDa poly-protein fusion construct comprising Ral2-TbH9-Ra35)
 CC linked to the Leishmania TSA (MAPS). (Updated on 29-AUG-2003 to
 CC standardise OS field)

XX
 SQ Sequence 2808 BP; 466 A; 889 C; 982 G; 471 T; 0 U; 0 Other;

Query Match 95.6%; Score 2186.8; DB 6; Length 2808;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	39	CATATGATCAACCATCAACATCAACAGCGCGCGTCCGATTAATTCAGCTGTCCAGGGT	98
Db	1	CATATGATCAACCATCAACATCAACAGCGCGCGTCCGATTAATTCAGCTGTCCAGGGT	60
Qy	99	GGGAGGAGTTCGGCATTCGGATCGGCGAGCGATGCGATCGGCGCGAGATCCGATCG	158
Db	61	GGGAGGAGTTCGGCATTCGGATCGGCGAGCGATGCGATCGGCGCGAGATCCGATCG	120
Qy	159	GGTGGGGGTCAACCAACCGTTTCATATCGGGGTACCGGCTTCTCGGCTTGGGTGTGTC	218
Db	121	GGTGGGGGTCAACCAACCGTTTCATATCGGGGTACCGGCTTCTCGGCTTGGGTGTGTC	180
Qy	219	GACACACACGACGCGACGAGTCCAGCGTGTGCGGAGCGCTCCGGCGGCAAGT	278
Db	181	GACACACACGACGCGACGAGTCCAGCGTGTGCGGAGCGCTCCGGCGGCAAGT	240
Qy	279	CTCGGCATCTCACCGGCGAGTATCACCGCGGTTCGACGGCGCTCCGATCAACTCGGC	338
Db	241	CTCGGCATCTCACCGGCGAGTATCACCGCGGTTCGACGGCGCTCCGATCAACTCGGC	300
Qy	339	ACCGGATGCGGAGCGGCTTAACGGGATCATCCCGGTGACGTATCTCGGTGACCTGG	398
Db	301	ACCGGATGCGGAGCGGCTTAACGGGATCATCCCGGTGACGTATCTCGGTGACCTGG	360
Qy	399	CAAAACAAAGTCGGCGGACGCGTACAGGGAAGTGCATTTGGCGGAGGACCCCGGC	458
Db	361	CAAAACAAAGTCGGCGGACGCGTACAGGGAAGTGCATTTGGCGGAGGACCCCGGC	420
Qy	459	GAATTCATGTGGATTTGGGGGCTTACACCGAGATCAACTCCCGGAGGATGTACGCC	518
Db	421	GAATTCATGTGGATTTGGGGGCTTACACCGAGATCAACTCCCGGAGGATGTACGCC	480
Qy	519	GGCCCGGTTTCGGCTCGCTGGTGGCGGCGCTCAGATGTGGACAGCTGGCGATGAC	578
Db	481	GGCCCGGTTTCGGCTCGCTGGTGGCGGCGCTCAGATGTGGACAGCTGGCGATGAC	540
Qy	579	CTGTGTTTCGGCGCGTTCGGCTTTCAGTGGTGGTCTGGGCTCTGACGGTGGGTCTGG	638
Db	541	CTGTGTTTCGGCGCGTTCGGCTTTCAGTGGTGGTCTGGGCTCTGACGGTGGGTCTGG	600
Qy	639	ATAGGTTCGTGGCGGGTCTGATGTGGCGGCGCTCCCGGTATGTGGCGTGGATGAGC	698
Db	601	ATAGGTTCGTGGCGGGTCTGATGTGGCGGCGCTCCCGGTATGTGGCGTGGATGAGC	660
Qy	699	GTCAACCGGGGAGGCGGAGCTGACCGCGCGCGAGTCCGGTTCGTCGCGCGGCTAC	758
Db	661	GTCAACCGGGGAGGCGGAGCTGACCGCGCGCGAGTCCGGTTCGTCGCGCGGCTAC	720
Qy	759	GAGACGGGTATGGGTGAGCGGTTCGGTGGCGGCGCTCCCGGTATGTGGCGTGGATGAGC	818
Db	721	GAGACGGGTATGGGTGAGCGGTTCGGTGGCGGCGCTCCCGGTATGTGGCGTGGATGAGC	780
Qy	819	ATGATTCATGATCGGACCAACTCTTTGGGGCAAAACACCCCGGCGATCGCGTCAACGAG	878
Db	781	ATGATTCATGATCGGACCAACTCTTTGGGGCAAAACACCCCGGCGATCGCGTCAACGAG	840

Qy	879	GCGGAATACGGCGAGATGGGCGCCCAAGACGCGCGCGGATGTTTGGCTACGCGCGCG	938
Db	841	GCGGAATACGGCGAGATGGGCGCCCAAGACGCGCGCGGATGTTTGGCTACGCGCGCG	900
Qy	939	ACGGCGACGCGACGCGGAGTTCGCTGCGGTTTCGAGGAGGCGCGGAGATGACACGCG	998
Db	901	ACGGCGACGCGACGCGGAGTTCGCTGCGGTTTCGAGGAGGCGCGGAGATGACACGCG	960
Qy	999	GGTGGGTCTCTTCAGCAGCGCGCGCTTCGAGGAGGCTTCGACACCGCGCGCGGCAAC	1058
Db	961	GGTGGGTCTCTTCAGCAGCGCGCGCTTCGAGGAGGCTTCGACACCGCGCGCGGCAAC	1020
Qy	1059	CAGTTGATGAACAATGTGCGCCAGCGGCTGCAACAGCTGGCGCCAGCCACGAGGGGACC	1118
Db	1021	CAGTTGATGAACAATGTGCGCCAGCGGCTGCAACAGCTGGCGCCAGCCACGAGGGGACC	1080
Qy	1119	ACGCGCTTCTTCCAAAGTGGGCGCTGTGGAAGACGCTCTCGCGCATCGCTCGCGGATC	1178
Db	1081	ACGCGCTTCTTCCAAAGTGGGCGCTGTGGAAGACGCTCTCGCGCATCGCTCGCGGATC	1140
Qy	1179	ACCAACATGGTGTGATGCGCAACCAACACATGTCGATGACAACTCGGGTGTGTCGATG	1238
Db	1141	ACCAACATGGTGTGATGCGCAACCAACACATGTCGATGACAACTCGGGTGTGTCGATG	1200
Qy	1239	ACCAACATCTTGGATGCTGATGTTGAAGGCTTGTCTCGCGCGCGCGCGCGCGCGTG	1298
Db	1201	ACCAACATCTTGGATGCTGATGTTGAAGGCTTGTCTCGCGCGCGCGCGCGCGCGTG	1260
Qy	1299	CAAAACCGCGCGCAAAACCGGCTTCGGGCGATGCTGTGGGACAGCTCGCTGGGTCT	1358
Db	1261	CAAAACCGCGCGCAAAACCGGCTTCGGGCGATGCTGTGGGACAGCTCGCTGGGTCT	1320
Qy	1359	TGGGTCTGGGCGGTGGGCTGCGCGCAACTTGGTGGGCGCGCGCTCGGTGGGTG	1418
Db	1321	TGGGTCTGGGCGGTGGGCTGCGCGCAACTTGGTGGGCGCGCGCTCGGTGGGTG	1380
Qy	1419	TCGTGTCGCGAGGCTTCGGCGCGCGCAACAGGACGTCAACCGCGCGCGCGCGCGCTG	1478
Db	1381	TCGTGTCGCGAGGCTTCGGCGCGCGCAACAGGACGTCAACCGCGCGCGCGCGCGCTG	1440
Qy	1479	CGGTGACCAAGCTGACAGCGCGCGGAGAGGCGCGGCGAGATGCTGGGCGGCTG	1538
Db	1441	CGGTGACCAAGCTGACAGCGCGCGGAGAGGCGCGGCGAGATGCTGGGCGGCTG	1500
Qy	1539	CCGTGGGCGAGATGGGCGCGCGCGCGCTTCGCGGCTCAGTGGTGTGCTGGGTTCG	1598
Db	1501	CCGTGGGCGAGATGGGCGCGCGCGCGCTTCGCGGCTCAGTGGTGTGCTGGGTTCG	1560
Qy	1599	CCGCGACCTATGTGATGCGCATCTTCGCGAGCGCGCGGATGTCGCCCGCGCGCTTG	1658
Db	1561	CCGCGACCTATGTGATGCGCATCTTCGCGAGCGCGCGGATGTCGCCCGCGCGCTTG	1620
Qy	1659	TCGACAGACCGGTTTCGCGCGCTTCGCGCGCTCGCGCTCGCGCTCGCGGATGTCGC	1718
Db	1621	TCGACAGACCGGTTTCGCGCGCTTCGCGCGCTCGCGCTCGCGCTCGCGGATGTCGC	1680
Qy	1719	CAAGTGGGCGACAGGTGGTCAACATCAACCAAACTGGGCTTACAAACCGCGTGGGC	1778
Db	1681	CAAGTGGGCGACAGGTGGTCAACATCAACCAAACTGGGCTTACAAACCGCGTGGGC	1740
Qy	1779	GCGGCGACCGGATCGATCGATCCCAAGGTGTGCTGCTGACCAACACCGCTGATC	1838
Db	1741	GCGGCGACCGGATCGATCGATCCCAAGGTGTGCTGCTGACCAACACCGCTGATC	1800
Qy	1839	GCGGCGCGCACCGACATCAATGGGTTTCAGGCTTCGGGCTTCGGCGCAAACTACGCGCTGAT	1898
Db	1801	GCGGCGCGCACCGACATCAATGGGTTTCAGGCTTCGGGCTTCGGCGCAAACTACGCGCTGAT	1860
Qy	1899	GTGGTTCGGGTATGACCGCACCGGATGTCGGGTGCTGCTGACGCTGCGGCTGCGGTGGC	1958
Db	1861	GTGGTTCGGGTATGACCGCACCGGATGTCGGGTGCTGCTGACGCTGCGGCTGCGGTGGC	1920
Qy	1959	CTGCGGTTCGCGCGGATTCGGTGGCGCGCTTCGGTGGGTTCGTCGAGCCCTCGCTGGGCTGGGC	2018

1059 CAGTTGATCAACAAATGTGCCCCCAGGCGCTGCAACAGCTGGCCACGACCCACGAGGGCACC 1118
1021 CAGTTGATGAACAATGTGCCCCCAGGCGCTGCAACAGCTGGCCACGACCCACGAGGGCACC 1080
1119 ACGCCCTTCTTCAAGCTGGGTGGCTGTGGAGACGGTCTCGCCCATCTGGTTCGCCGATC 1178
1081 ACGCCTTCTTCAAGCTGGGTGGCTGTGGAGACGGTCTCGCCCATCTGGTTCGCCGATC 1140
1179 AGCAACATGGTTCGATGGCCCAACAAACACATGTGATGACCACTCGGGTGTGTCGATG 1238
1141 AGCAACATGGTTCGATGGCCCAACAAACACATGTGATGACCACTCGGGTGTGTCGATG 1200
1239 ACCAACACCTTGGATCGATGTTGAAGGCTTTGCTCGGCGGGCGCCCGCAGGCGGTG 1298
1201 ACCAACACCTTGGATCGATGTTGAAGGCTTTGCTCGGCGGGCGCCCGCAGGCGGTG 1260
1299 CAAACCGGCGGCAAAACGGGGTCCGGCGGATGAGTCTGCTGGGACAGTCTCGCTGGTTC 1358
1261 CAAACCGGCGGCAAAACGGGGTCCGGCGGATGAGTCTGCTGGGACAGTCTCGCTGGTTC 1320
1359 TCGGCTCTGGGCGGTGGGCTGGCCCAACTTGGTGGGCGGCGCTCGGTCGGTTCGTTG 1418
1321 TCGGCTCTGGGCGGTGGGCTGGCCCAACTTGGTGGGCGGCGCTCGGTCGGTTCGTTG 1380
1419 TCGGTCGCGCAGGCTGGGCGGCGGCAACAGGACAGTCAACCGGCGGCGGCGGCGCTG 1478
1381 TCGGTCGCGCAGGCTGGGCGGCGGCAACAGGACAGTCAACCGGCGGCGGCGGCGCTG 1440
1479 CCGCTGACAGCCTGACAGCGCGCGGCAAGAGGGCCCGGACAGTCTGGGCGGCGCTG 1538
1441 CCGCTGACAGCCTGACAGCGCGCGGCAAGAGGGCCCGGACAGTCTGGGCGGCGCTG 1500
1539 CCGGTGGGCGAGATGGGCGGCGGCGGCTGGTGGGCTCAAGTGTGCTGCGGTTCCG 1598
1501 CCGGTGGGCGAGATGGGCGGCGGCGGCTGGTGGGCTCAAGTGTGCTGCGGTTCCG 1560
1599 CCGGACCCCTATGATGCGCCATCTCCGACGCGCGGATATCGCCCGGCGGCGCTG 1658
1561 CCGGACCCCTATGATGCGCCATCTCCGACGCGCGGATATCGCCCGGCGGCGCTG 1620
1659 TCGCAGGACCCGTTCCGCGCTTCCCGCGCTGCGCCCTCGACCGCTCGCGGATGTCGC 1718
1621 TCGCAGGACCCGTTCCGCGCTTCCCGCGCTGCGCCCTCGACCGCTCGCGGATGTCGC 1680
1719 CAGGTGGGCGCACAGGTGGTCAACATCAACCAAACTGGGCTACCAACACGCGCTGGG 1778
1681 CAGGTGGGCGCACAGGTGGTCAACATCAACCAAACTGGGCTACCAACACGCGCTGGG 1740
1779 GCCGGGACCGGCATCGTCATCGATCCCAACGCTGCTGCTGACCAACACGCTGATC 1838
1741 GCCGGGACCGGCATCGTCATCGATCCCAACGCTGCTGCTGACCAACACGCTGATC 1800
1839 GCGGCGCCACCGACATCAATGGCTTCAGCGTGGGCTCCGGCCAAACCTACGCGCTCGAT 1898
1801 GCGGCGCCACCGACATCAATGGCTTCAGCGTGGGCTCCGGCCAAACCTACGCGCTCGAT 1860
1899 GTGTCGGGTATGACGCGACCAAGATGTCGGTCTGAGTCTGAGTGGCGGTCGGTGGC 1958
1861 GTGTCGGGTATGACGCGACCAAGATGTCGGTCTGAGTCTGAGTGGCGGTCGGTGGC 1920
1959 CTGCGCTCGGCGGATCGGTGGCGGCTCGCGGTTGGTGGAGCCGCTCGTTCGCGATGGC 2018
1921 CTGCGCTCGGCGGATCGGTGGCGGCTCGCGGTTGGTGGAGCCGCTCGTTCGCGATGGC 1980
2019 AACAGCGTGGGACGGCGGAAACCGCCGCTGGTGGCTGGAGGCTGTGCGCTCGC 2078
1981 AACAGCGTGGGACGGCGGAAACCGCCGCTGGTGGCTGGAGGCTGTGCGCTCGC 2040
2079 CAAACCGTGGGCGGCTCGGATCGCTGACCGGTCGGAAGACATTTGAACGGGTGATC 2138
2041 CAAACCGTGGGCGGCTCGGATCGCTGACCGGTCGGAAGACATTTGAACGGGTGATC 2100
2139 CAGTTGATGCGCGCATCGACCCCGGTGATTCGGGCGGCGGCTCGTCAACGCGCTAGGA 2198

Db 2101 CAGTTGATGCGCGCATCCAGCCCGGTGATTCGGGCGGCGGCTCGTCAACGCGCTAGGA 2160
QY 2199 CAGGTGGTGGTATGAACACGCGCGGTCC 2228
Db 2161 CAGGTGGTGGTATGAACACGCGCGGTCC 2190
RESULT 11
ADA26363
ID ADA26363 standard; DNA; 3060 BP.
XX ADA26363;
AC ADA26363;
XX
DT 20-NOV-2003 (first entry)
XX
DE M. bovis MTB72F and 85b complex antigen (fusion MTB103F) DNA.
XX
KW ds: gene; fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
XX tuberculosis; tuberculostatic; gene therapy; vaccine; MTB72F; 85B.
OS Chimeric.
XX Mycobacterium bovis.
XX
PH Key Location/Qualifiers
FT CDS 4..3054
FT /*tag= a
FT /product= "MTB72F and 85b complex (fusion MTB103F) "
XX
PN WO2003070187-A2.
XX
PD 28-AUG-2003.
XX
PF 18-FEB-2003; 2003WO-US004903.
XX
PR 15-FEB-2002; 2002US-0357351P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Guderian J, Reed S;
XX
DR WPI; 2003-697554/66.
XX
PS P-PSDB; ADA26370.
XX
CC New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
CC and MTB39 antigens, with or without the MTB85A antigen, from a
CC Mycobacterium species, useful for diagnosing, preventing and/or treating
CC tuberculosis infection.
XX
PS Claim 84; Fig 11; 112pp; English.
XX
CC The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polypeptide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The
CC present sequence is used in the exemplification of the invention.
XX
SQ Sequence 3060 BP; 503 A; 977 C; 1070 G; 510 T; 0 U; 0 Other;
Query Match 95.6%; Score 2186.8; DB 8; Length 3060;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 39 CATATGATCACCATCACCATCACAGCGCGCTCGGATAACTTCCAGTGTCCAGGGT 98
Db 1 CATATGATCACCATCACCATCACAGCGCGCTCGGATAACTTCCAGTGTCCAGGGT 60
QY GGGCAGGATTCGCCATTCGATCGGCGAGCGGATCGGCGGCGAGATCCGATCG 158
Db 61 GGGCAGGATTCGCCATTCGATCGGCGAGCGGATCGGCGGCGAGATCCGATCG 120

AC ADA26362;
XX 20-NOV-2003 (first entry)
XX Mycobacterium MTB72F-hTCC#1 (fusion MTB102tm2F) protein, encoding DNA.
XX ds; gene; fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW tuberculosis; tuberculostatic; gene therapy; vaccine.
XX Chimeric.
OS Mycobacterium sp.
XX Key Location/Qualifiers
FT CDS 4..3072
FT /*tag= a
FT /product= "MTB72F-hTCC#1 (fusion MTB102tm2F) protein"
XX
PN WO2003070187-A2.
XX 28-AUG-2003.
XX 18-FEB-2003; 2003WO-US004903.
XX 15-FEB-2002; 2002US-0357351P.
XX (CORI-) CORIXA CORP.
XX Skeiky Y, Gudexian J, Reed S;
XX WPI; 2003-697554/66.
XX P-PSDB; ADA26369.
XX
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
XX and MTB39 antigens, with or without the MTB85A antigen, from a
XX Mycobacterium species, useful for diagnosing, preventing and/or treating
XX tuberculosis infection.
XX
XX Claim 84; Fig 10; 112pp; English.
XX
XX The invention relates to a novel nucleic acid encoding a fusion
XX polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
XX MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
XX A polypeptide of the invention has tuberculostatic activity. A
XX polynucleotide of the invention may have a use in gene therapy, and as a
XX vaccine. The methods and compositions of the present invention are useful
XX for diagnosing, preventing and/or treating tuberculosis infection. The
XX present sequence is used in the exemplification of the invention.
XX
XX Sequence 3104 BP; 510 A; 968 C; 1104 G; 522 T; 0 U; 0 Other;
XX
XX Query Match 95.6%; Score 2186.8; DB 8; Length 3104;
XX Best Local Similarity 99.9%; Pred. No. 0;
XX Matches 2186; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX 39 CATATGATCACCATCACCATCACCACGGCGCGTCCGATAAATCCAGCTGTCCAGGTT 98
XX 1 CATATGATCACCATCACCATCACCACGGCGCGTCCGATAAATCCAGCTGTCCAGGTT 60
XX
XX 99 GGGCAGGGATCCGATCCGATCCGCGAGCGATCCGATCCGCGCCAGATCCGATCCG 158
XX 61 GGGCAGGGATCCGATCCGATCCGCGAGCGATCCGATCCGCGCCAGATCCGATCCG 120
XX
XX 159 GGTGGGGGGTCAACCCAGCTTATATCGGGCCCTACCGCCCTTCTCGGCTTGGGTGTC 218
XX 121 GGTGGGGGGTCAACCCAGCTTATATCGGGCCCTACCGCCCTTCTCGGCTTGGGTGTC 180
XX
XX 219 GACAAACGGCGAAGCGGACGAGTCCAGCGTGTGCGGAGCGCTCCGCGGCAAGT 278
XX 181 GACAAACGGCGAAGCGGACGAGTCCAGCGTGTGCGGAGCGCTCCGCGGCAAGT 240
XX
XX 279 CTGGGATCTCCACCGCGAGCGTATCACCAGCGGTGACCGCGCTCCGATCAACTCGGCC 338
XX 241 CTGGGATCTCCACCGCGAGCGTATCACCAGCGGTGACCGCGCTCCGATCAACTCGGCC 300

339 ACCGGATGGCGGACGCGCTTAACGGGCATCATCCCGGTGACGTATCTCGGTGACCTGG 398
|
|
|
301 ACCGGATGGCGGACGCGCTTAACGGGCATCATCCCGGTGACGTATCTCGGTGACCTGG 360
|
|
|
399 CAAACCAAGTCGGGCGGACGCGTACAGGGAACGTGACATTGGCGGAGGACCCCGGCC 458
|
|
|
361 CAAACCAAGTCGGGCGGACGCGTACAGGGAACGTGACATTGGCGGAGGACCCCGGCC 420
|
|
|
459 GAATTCATGGTGGATTTCGGGGCGTTACACGGGAGATCACTCGGAGGATGTAGCC 518
|
|
|
421 GAATTCATGGTGGATTTCGGGGCGTTACACGGGAGATCACTCGGAGGATGTAGCC 480
|
|
|
519 GCGCCGGGTTCGGCCCTCGCTCGTGGCCCGCGCTCAGATGTGGGACAGCGTGGCGAGTAC 578
|
|
|
481 GCGCCGGGTTCGGCCCTCGCTCGTGGCCCGCGCTCAGATGTGGGACAGCGTGGCGAGTAC 540
|
|
|
579 CTGTTTTCGGCGCGTTCGGCGTTTCAGTTCGTGGTCTGGGTCTGACGGTGGGTCTGG 638
|
|
|
541 CTGTTTTCGGCGCGTTCGGCGTTTCAGTTCGTGGTCTGGGTCTGACGGTGGGTCTGG 600
|
|
|
639 ATAGGTTTCGTGGCGGTCTGATGTGGCGCGCTCGCGCTATGTGGCGTGGATGAGC 698
|
|
|
601 ATAGGTTTCGTGGCGGTCTGATGTGGCGCGCTCGCGCTATGTGGCGTGGATGAGC 660
|
|
|
699 GTACCGCGGCGAGGCGGCTGACGCGCGCGCGCTCGCGTTCGGGTTCGTGGCGCGCTAC 758
|
|
|
661 GTACCGCGGCGAGGCGGCTGACGCGCGCGCGCTCGCGTTCGGGTTCGTGGCGCGCTAC 720
|
|
|
759 GAGCGCGGTATGGCTGACGCGCGCGCGCTGATCGCGGAGAACCGTCTGAACTG 818
|
|
|
721 GAGCGCGGTATGGCTGACGCGCGCGCGCTGATCGCGGAGAACCGTCTGAACTG 780
|
|
|
819 ATGATTCGTAGGACCAACCTCTTGGGCAAAACACCCCGCGGATCGCGGTCAACGAG 878
|
|
|
781 ATGATTCGTAGGACCAACCTCTTGGGCAAAACACCCCGCGGATCGCGGTCAACGAG 840
|
|
|
879 GCGGATACGGGAGATGTGGCCCAAGACCCCGCGGATGTTGGCTAGCCCGCGCG 938
|
|
|
841 GCGGATACGGGAGATGTGGCCCAAGACCCCGCGGATGTTGGCTAGCCCGCGCG 900
|
|
|
939 ACCGCGACGGCGAGCGGCGCTGCTGCGCTTCAGGAGGCGCGGAGATGACACAGCGCG 998
|
|
|
901 ACCGCGACGGCGAGCGGCGCTGCTGCGCTTCAGGAGGCGCGGAGATGACACAGCGCG 960
|
|
|
999 GGTGGGTCTCTGAGAGCGCGCGCGTTCAGGAGGCTTCGACACCGCGCGCGGCAAC 1058
|
|
|
961 GGTGGGTCTCTGAGAGCGCGCGCGTTCAGGAGGCTTCGACACCGCGCGCGGCAAC 1020
|
|
|
1059 CAGTTGATGAACAATGTGCGCCAGCGCTGCAACAGCTGGCCAGCCACACAGCGGCAAC 1118
|
|
|
1021 CAGTTGATGAACAATGTGCGCCAGCGCTGCAACAGCTGGCCAGCCACACAGCGGCAAC 1080
|
|
|
1119 ACAGCTTCTTCAAGCTGGGTGGCTGTGGAAGACGGTTCGCGCGATTCGGTTCGCCGATC 1178
|
|
|
1081 ACAGCTTCTTCAAGCTGGGTGGCTGTGGAAGACGGTTCGCGCGATTCGGTTCGCCGATC 1140
|
|
|
1179 AGCAACATGGTTCGATGGCCCAACACACATGTCGATGACCACTCGGTGTGTCATG 1238
|
|
|
1141 AGCAACATGGTTCGATGGCCCAACACACATGTCGATGACCACTCGGTGTGTCATG 1200
|
|
|
1239 ACCAACACCTTTGAGCTCGATTTGAAGGCTTTTCTCCGGCGGCGCCCGCCAGCGCGTG 1298
|
|
|
1201 ACCAACACCTTTGAGCTCGATTTGAAGGCTTTTCTCCGGCGGCGCCCGCCAGCGCGTG 1260
|
|
|
1299 CAAACCGCGCGCAAAACGGGTTCGGCGGATGAGCTCGTGGGAGCTCGTGGGTCT 1358
|
|
|
1261 CAAACCGCGCGCAAAACGGGTTCGGCGGATGAGCTCGTGGGAGCTCGTGGGTCT 1320
|
|
|
1359 TCGGGTCTGGGCGGTGGGTCGCCCAACTTTGGGTTCGGCGGCTCGGTTCGGTTCG 1418
|
|
|
1321 TCGGGTCTGGGCGGTGGGTCGCCCAACTTTGGGTTCGGCGGCTCGGTTCGGTTCG 1380

QY 1419 TCGGTGCGCAGGCTTGGGCGGCGCAACAGGAGTCAACCCGCGCGGGGCGT 1478
Db 1381 TCGGTGCCGAGGCTTGGGCGGCGCAACAGGAGTCAACCCGCGCGGGGCGT 1440
QY 1479 CCGCTACACAGCTGACACGCGCGGAAAGAGGCGCGGAGATGCTGGCGGGGCTG 1538
Db 1441 CCGCTACACAGCTGACACGCGCGGAAAGAGGCGCGGAGATGCTGGCGGGCTG 1500
QY 1539 CCGGTGGGCGAGATGGGCGGCGAGGCGGCTGGTGGCTCAGTGGTGGTGGTTCGG 1598
Db 1501 CCGGTGGGCGAGATGGGCGGCGAGGCGGCTGGTGGCTCAGTGGTGGTTCGG 1560
QY 1599 CCGGACCTATGCTGATGCGGATCTCCGCGAGCGGCGGATCGCCCGCGGCTTG 1658
Db 1561 CCGGACCTATGCTGATGCGGATCTCCGCGAGCGGCGGATCGCCCGCGGCTTG 1620
QY 1659 TCGAGGACCGGTTCCGCACTTCCCGCGCTGCCCTCGACCGGTCGCGATGTCG 1718
Db 1621 TCGAGGACCGGTTCCGCACTTCCCGCGCTGCCCTCGACCGGTCGCGATGTCG 1680
QY 1719 CAAAGTGGGCGCAGAGTGTCAACATCAACCAACAGGCTGACCAACCAACGCTGATC 1778
Db 1691 CAAAGTGGGCGCAGAGTGTCAACATCAACCAACAGGCTGACCAACCAACGCTGATC 1740
QY 1779 GCGGAGCGGCTGCTCATCGATCCCAACGCTGCTGCTGACCAACCAACGCTGATC 1838
Db 1741 GCGGAGCGGCTGCTCATCGATCCCAACGCTGCTGCTGACCAACCAACGCTGATC 1800
QY 1839 GCGGCGCACGACATCAATGGTTGAGCTGCGCTCGGCTCGGCAACCTACCGGCTGAT 1898
Db 1801 GCGGCGCACGACATCAATGGTTGAGCTGCGCTCGGCTCGGCAACCTACCGGCTGAT 1860
QY 1899 GTGGTGGGCTATGACCGCACCGAGGATGCGGGTGTGCTGAGCTGCGGCTGCGGCTG 1958
Db 1861 GTGGTGGGCTATGACCGCACCGAGGATGCGGGTGTGCTGAGCTGCGGCTGCGGCTG 1920
QY 1959 CTGCTGCGGCGGATGCGGTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 2018
Db 1921 CTGCTGCGGCGGATGCGGTGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1980
QY 2019 AACAGCGGTGGGCGGCGGAAAGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 2078
Db 1981 AACAGCGGTGGGCGGCGGAAAGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 2040
QY 2079 CAAACCGTGCGGCTGCGGATGCGTGCACCGGTCGCGAAGAGACATTTGAACGGTTGATC 2138
Db 2041 CAAACCGTGCGGCTGCGGATGCGTGCACCGGTCGCGAAGAGACATTTGAACGGTTGATC 2100
QY 2139 CAGTTGATGCGGCTGCGGATGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 2198
Db 2101 CAGTTGATGCGGCTGCGGATGCGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 2160
QY 2199 CAGGTGGTGGTATGAACACGCGGCGGCTG 2228
Db 2161 CAGGTGGTGGTATGAACACGCGGCGGCTG 2190

RESULT 13
ADA26361
ID ADA26361 standard; DNA; 3474 BP.
XX AC ADA26361;
XX AC ADA26361;
DT 20-NOV-2003 (first entry)
XX Mycobacterium MTB72F-mTCC#2 (fusion MTB114F) protein encoding DNA.
DE ds; gene; fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW tuberculosis; tuberculostatic; gene therapy; vaccine.
XX Chimeric.
OS Mycobacterium sp.
XX

PH Key Location/Qualifiers
FT CDS 4..3468
FT /*tag= a
FT /product= "MTB72F-mTCC#2 (fusion MTB114F) protein"
XX WO2003070187-A2.
XX 28-AUG-2003.
XX 18-FEB-2003; 2003WO-US004903.
XX 15-FEB-2002; 2002US-0357351P.
XX (CORI-) CORIXA CORP.
XX Skeiky Y, Guderian J, Reed S;
XX WPI; 2003-697554/66.
XX P-PSDB; ADA26368.
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
XX and MTB39 antigens, with or without the MTB85A antigen, from a
XX Mycobacterium species, useful for diagnosing, preventing and/or treating
XX tuberculosis infection.
XX Claim 84; Fig 9; 112pp; English.
XX The invention relates to a novel nucleic acid encoding a fusion
XX polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
XX MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
XX A polypeptide of the invention has tuberculostatic activity. A
XX polynucleotide of the invention may have a use in gene therapy, and as a
XX vaccine. The methods and compositions of the present invention are useful
XX for diagnosing, preventing and/or treating tuberculosis infection. The
XX present sequence is used in the exemplification of the invention.
XX Sequence 3474 BP; 548 A; 1131 C; 1252 G; 543 T; 0 U; 0 Other;

Query Match 95.6%; Score 2186.8; DB 8; Length 3474;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 39 CATATGATCATCATCACCATCAACGCGCGCGCTCCGATTAACCTCCAGCTGTCACGGGT 98
Db 1 CATATGATCATCATCACCATCAACGCGCGCGCTCCGATTAACCTCCAGCTGTCACGGGT 60
QY 99 GGGCAGGAGATTCCGCTATCCGATCGGCGAGCGATGCGATCGGCGCGCAGATCCGATCG 158
Db 61 GGGCAGGAGATTCCGCTATCCGATCGGCGAGCGATGCGATCGGCGCGCAGATCCGATCG 120
QY 159 GGTGGGGGTCAACCAACCGTTTCATCGGCGCTTACCGCTTCTCTCGGCTGGGTGTC 218
Db 121 GGTGGGGGTCAACCAACCGTTTCATCGGCGCTTACCGCTTCTCTCGGCTGGGTGTC 180
QY 219 GACACACGCGCAACGCGCGAGTCCAGCGCTGTCGCGAGCGCTCCGCGCGCAAGT 278
Db 181 GACACACGCGCAACGCGCGAGTCCAGCGCTGTCGCGAGCGCTCCGCGCGCAAGT 240
QY 279 CTCGGCATCTCCACCGCGAGCTGATCAACCGCGTTCGACGCGCTCCGATCAATCCGCC 338
Db 241 CTCGGCATCTCCACCGCGAGCTGATCAACCGCGTTCGACGCGCTCCGATCAATCCGCC 300
QY 339 ACCGGATGGGAGCGGCTTAAACGGCTATCCCGGTGACGCTCATCTCGGTGACCTGG 398
Db 301 ACCGGATGGGAGCGGCTTAAACGGCTATCCCGGTGACGCTCATCTCGGTGACCTGG 360
QY 399 CAAACCAAGTCCGGCGCGCACGCTACAGGAACTGACATTTGGCCGAGGAGCCCCCGCC 458
Db 361 CAAACCAAGTCCGGCGCGCACGCTACAGGAACTGACATTTGGCCGAGGAGCCCCCGCC 420
QY 459 GAATTCATGTTGGATTTTCGGGGGCTTACCGCGAGATCAACTCCGCGAGATGATGCC 518
Db 421 GAATTCATGTTGGATTTTCGGGGGCTTACCGCGAGATCAACTCCGCGAGATGATGCC 480

QY 519 GGCCTGGGTTCCGCTCCGCTGGTGGCCGCGCTCAGATGTGGACACGCTGGCGAGTGAC 578
DB 481 GGCCTGGGTTCCGCTCCGCTGGTGGCCGCGCTCAGATGTGGACACGCTGGCGAGTGAC 540
QY 579 CTGTTTTCGGCCGCTGGGCTGTTTCACTCCGCTGGTGGGCTGTGAGCGTGGGGTGGTGG 638
DB 541 CTGTTTTCGGCCGCTGGGCTGTTTCACTCCGCTGGTGGGCTGTGAGCGTGGGGTGGTGG 600
QY 639 ATAGGTTTCGTCGGCGGGTCTGATGTTGGCGCGCGCTCGCGTATGTGGCGTGGATGAGC 698
DB 601 ATAGGTTTCGTCGGCGGGTCTGATGTTGGCGCGCGCTCGCGTATGTGGCGTGGATGAGC 660
QY 699 GTACCGCGGGGAGCGCGAGCTGACCGCGCGCGCGCTCGGGTGTGTCGGCGCGCGCTAC 758
DB 661 GTACCGCGGGGAGCGCGAGCTGACCGCGCGCGCGCTCGGGTGTGTCGGCGCGCGCTAC 720
QY 759 GAGACGGGCTATGGGCTGACGGTCCCGCGCGCTGATCCCGAGAACCGCTGTAACCTG 818
DB 721 GAGACGGGCTATGGGCTGACGGTCCCGCGCGCTGATCCCGAGAACCGCTGTAACCTG 780
QY 819 ATGATTTCTGATAGCGACCAACCTTTGGGGGCAAAACACCCCGCGGATCGCGTCAACGAG 878
DB 781 ATGATTTCTGATAGCGACCAACCTTTGGGGGCAAAACACCCCGCGGATCGCGTCAACGAG 840
QY 879 GCCGAATACGGCGAGATGTGGGCGCAAGACCGCGCGCGCTGTTGGCTACGCGCGGGCG 938
DB 841 GCCGAATACGGCGAGATGTGGGCGCAAGACCGCGCGCGCTGTTGGCTACGCGCGGGCG 900
QY 939 ACGCCGACGCGACGCGGCAAGTTGCTCCGTTTGAAGAGCGCGCGGAGATGACACGCGG 998
DB 901 ACGCCGACGCGACGCGGCAAGTTGCTCCGTTTGAAGAGCGCGCGGAGATGACACGCGG 960
QY 999 GGTGGGCTCTCGAGCAGGCGCGCGCTCGAGAGGCTCCGACACCGCGCGCGGAGAC 1058
DB 961 GGTGGGCTCTCGAGCAGGCGCGCGCTCGAGAGGCTCCGACACCGCGCGCGGAGAC 1020
QY 1059 CAGTTGATGAACAAATGTGCCCGACAGCGCTGCAACAGCTGGCGCGCGCGCGAGGCGACC 1118
DB 1021 CAGTTGATGAACAAATGTGCCCGACAGCGCTGCAACAGCTGGCGCGCGCGAGGCGACC 1080
QY 1119 ACGCTTCTTCAAGCTGGTGGCTGCTGGAAGACGCTCTCGCGGATCGGTCGCGATC 1178
DB 1081 ACGCTTCTTCAAGCTGGTGGCTGCTGGAAGACGCTCTCGCGGATCGGTCGCGATC 1140
QY 1179 AGCAACATGGTTCGATGGCGCAACACCACTGTCGATGACCAACTCGGCTGTGTCGATG 1238
DB 1141 AGCAACATGGTTCGATGGCGCAACACCACTGTCGATGACCAACTCGGCTGTGTCGATG 1200
QY 1239 ACCAACCTTGAAGCTCGATGTTGAAGGCTTTGCTCCGCGCGCGCGCGCGCGCGCTG 1298
DB 1201 ACCAACCTTGAAGCTCGATGTTGAAGGCTTTGCTCCGCGCGCGCGCGCGCGCGCTG 1260
QY 1299 CAACCCGCGGCAAAACGCGGCTCCGCGCGATGAGCTCGCTGGCGAGCTCGCTGGGTTCT 1358
DB 1261 CAACCCGCGGCAAAACGCGGCTCCGCGCGATGAGCTCGCTGGCGAGCTCGCTGGGTTCT 1320
QY 1359 TCGGGTCTGGGCGGTGGGGTGGCGCGCAACTTGGGTCGCGCGCTCGGTCGCTGTTG 1418
DB 1321 TCGGGTCTGGGCGGTGGGGTGGCGCGCAACTTGGGTCGCGCGCTCGGTCGCTGTTG 1380
QY 1419 TCGGTGCGGAGGCTGGGCGCGGCGCAACACGAGCTACCCCGCGCGCGCGCGCGCTG 1478
DB 1381 TCGGTGCGGAGGCTGGGCGCGGCGCAACACGAGCTACCCCGCGCGCGCGCGCGCTG 1440
QY 1479 CCGCTGACGAGCTGACACGCGCGCGGAAAGAGGCGCGCGGAGAGTGTGGCGCGGCTG 1538
DB 1441 CCGCTGACGAGCTGACACGCGCGCGGAAAGAGGCGCGCGGAGAGTGTGGCGCGGCTG 1500
QY 1539 CCGGTGGGCGAGTGGGCGCGGCGCGGCTGTGGGCTCAGTGTGTGCTGCTGCTGCTG 1598
DB 1501 CCGGTGGGCGAGTGGGCGCGGCGCGGCTGTGGGCTCAGTGTGTGCTGCTGCTGCTGCTG 1560

QY 1599 CCGGACCCCTATGTGATGTCGCGCATTTCTCCGACGCGCGGATATCGCCCCCGGCTTG 1658
DB 1561 CCGGACCCCTATGTGATGTCGCGCATTTCTCCGACGCGCGGATATCGCCCCCGGCTTG 1620
QY 1659 TCGCAGGACCCGCTTCCGCGACTTTCCCGCGCTGCCCTTCGACCCGCTCGCGATGCTGGCC 1718
DB 1621 TCGCAGGACCCGCTTCCGCGACTTTCCCGCGCTGCCCTTCGACCCGCTCGCGATGCTGGCC 1680
QY 1719 CAAATGGGGCCACAGGTGGTCAACATCAACACCAAACTGGGCTACAAACACCGCGTGGCC 1778
DB 1681 CAAATGGGGCCACAGGTGGTCAACATCAACACCAAACTGGGCTACAAACACCGCGTGGCC 1740
QY 1779 GCGGGACCGGATCGTCATCGATCCCAACCGGTGTGCTGACCAACCAACACCGTGTATC 1838
DB 1741 GCGGGACCGGATCGTCATCGATCCCAACCGGTGTGCTGACCAACCAACACCGTGTATC 1800
QY 1839 GCGGGCCGACCGACATCAATCGCTTCAGCTCGCTCCGCCAAACCTACGGGCTCGAT 1898
DB 1801 GCGGGCCGACCGACATCAATCGCTTCAGCTCGCTCCGCCAAACCTACGGGCTCGAT 1860
QY 1899 GTGCTCGGATGACCGACCGGATGTCCGGTGTGCTGACGCTCGCGGTGCGGCTGGC 1958
DB 1861 GTGCTCGGATGACCGACCGGATGTCCGGTGTGCTGACGCTCGCGGTGCGGCTGGC 1920
QY 1959 CTGCGCTCGGCGGCGATCGGTGGGCGCTCGCGTGTGCTGACCGCGCTCGATGCGG 2018
DB 1921 CTGCGCTCGGCGGCGATCGGTGGGCGCTCGCGTGTGCTGACCGCGCTCGATGCGG 1980
QY 2019 AACCGCTGGGCGGCGGAAACCGCGCTCGCTGCTGACGCTGCGGCTGGCTGGC 2078
DB 1981 AACCGCTGGGCGGCGGAAACCGCGCTCGCTGCTGACGCTGCGGCTGGCTGGC 2040
QY 2079 CAAACCGCTGAGGCTCGGATTCGCTGACCGCTGCGAAGAGACATTAACCGGTTGATC 2138
DB 2041 CAAACCGCTGAGGCTCGGATTCGCTGACCGCTGCGAAGAGACATTAACCGGTTGATC 2100
QY 2139 CAGTTTCGATCGCGGATCCAGCCCGGATTTGCGGGCGGCGCGCTCGTCAACCGCTAGGA 2198
DB 2101 CAGTTTCGATCGCGGATCCAGCCCGGATTTGCGGGCGGCGCGCTCGTCAACCGCTAGGA 2160
QY 2199 CAGGTGGTTCGATCAACACCGCGCGCTCC 2228
DB 2161 CAGGTGGTTCGATCAACACCGCGCGCTCC 2190

RESULT 14

AAD47084

ID AAD47084 standard; DNA; 2190 BP.

XX AAD47084;

XX AC AAD47084;

XX 29-AUG-2003 (revised)

DT 27-JAN-2003 (first entry)

XX 29-AUG-2003 (revised)

XX Mycobacterium sp. MTB72FmutSA fusion protein encoding DNA.

XX Vaccine; immunity; diagnostic agent; gene therapy; TBH9; antigen;

XX Ra35MutSA; Ra12; MTB72MutSA; chimeric; gene; ds.

XX Mycobacterium sp.

OS Mycobacterium tuberculosis.

XX Chimeric.

XX Key Location/Qualifiers

FH CDS 1..2190

FT /*tag= a

FT /product= "MTB72FmutSA fusion protein"

XX WO200272792-A2.

XX 19-SEP-2002.

XX 13-MAR-2002; 2002WO-US008223.

Db 1561 CGACCTATGTATGCCGCAATTCTCCGGCAGCCGCGATATCGCCCGCGCCCTTGTG 1620
 QY 1662 CAGGACCGGTTCCCGCGCTGCCCGCTCCGACCGCTCCGCGATGTCGCCCAA 1721
 Db 1621 CAGGACCGGTTCCCGCGCTGCCCGCTCCGACCGCTCCGCGATGTCGCCCAA 1680
 QY 1722 GTGGGCGCACAGGTGTTCAACATCAACCAAACTGGGCTACAAACACCGCGTGGGCGC 1781
 Db 1681 GTGGGCGCACAGGTGTTCAACATCAACCAAACTGGGCTACAAACACCGCGTGGGCGC 1740
 QY 1782 GGGACCGGCATGTCATCGATCCCAACGGTGTGCTGTGACCAACCAACCGATCGG 1841
 Db 1741 GGGACCGGCATGTCATCGATCCCAACGGTGTGCTGTGACCAACCAACCGATCGG 1800
 QY 1842 GGGACCGGCATCAATGCTTTCAGCGTCCGCTCCGCGCAAACTACGGCGTGTG 1901
 Db 1801 GGGACCGGCATCAATGCTTTCAGCGTCCGCTCCGCGCAAACTACGGCGTGTG 1860
 QY 1902 GTCGGTATGACCGACCCAGATGTGCGGTGTGCTGACGCTGCGCGTGGGCTG 1961
 Db 1861 GTCGGTATGACCGACCCAGATGTGCGGTGTGCTGACGCTGCGCGTGGGCTG 1920
 QY 1962 CCGTGGCGGCGATCGGTGCGCGGTGCGGTGTGCTGACGCGCTGCGGATGGGCAAC 2021
 Db 1921 CCGTGGCGGCGATCGGTGCGCGGTGCGGTGTGCTGACGCGCTGCGGATGGGCAAC 1980
 QY 2022 AGCGTGGCGGCGGCGGAAACCCCGCTGCGGTGTGCTGACGCGGTGCGGCTGCGGCAAC 2081
 Db 1981 AGCGTGGCGGCGGCGGAAACCCCGCTGCGGTGTGCTGACGCGGTGCGGCTGCGGCAAC 2040
 QY 2082 ACCGTGCAAGCGTGGGATTCCTGCTGACCGGTGCGGAGAGACATGACGGGTGATCCAG 2141
 Db 2041 ACCGTGCAAGCGTGGGATTCCTGCTGACCGGTGCGGAGAGACATGACGGGTGATCCAG 2100
 QY 2142 TTCGATGCGCGATCCAGCCCGGTGATTCGGCGGCGGCGCGTCTGCTCAACGGCTAGGACAG 2201
 Db 2101 TTCGATGCGCGATCCAGCCCGGTGATTCGGCGGCGGCGCGTCTGCTCAACGGCTAGGACAG 2160
 QY 2202 GTGGTGGGTATGACACGGCGCGCTCTAG 2231
 Db 2161 GTGGTGGGTATGACACGGCGCGCTCTAG 2190

RESULT 15
 AAD28343

ID AAD28343 standard; DNA; 2190 BP.

AC AAD28343;

DT 22-APR-2002 (first entry)

XX Mycobacterium species MTB72FmutSA fusion protein encoding DNA.

KW Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB72FmutSA; Ra12-TbH9-Ra35MutSA;
 KW mutant; mutin; ds.

OS Mycobacterium sp.

PH Key Location/Qualifiers

FT CDS 1..2190

FT /*tag= a

FT /product= "MTB72FmutSA fusion protein"

FT 22..417

FT misc_feature

FT /note= "Ra12 DNA fragment"

FT 424..1596

FT misc_feature

FT /*tag= c

FT /note= "TBH9FL DNA fragment"

FT 1603..2187

FT misc_feature

FT /*tag= d

FT /note= "Ra35 DNA fragment"

FT mutation

FT /*tag= e

XX WO200198460-A2.

XX 27-DEC-2001.

XX 20-JUN-2001; 2001WO-US019959.

XX 20-JUN-2000; 2000US-00597796.

XX 01-FEB-2001; 2001US-0265737P.

XX (CORI-) CORIXA CORP.

XX Skeiky Y, Reed S, Alderson M;

XX WPI; 2002-147798/19.

XX P-PSDB; AAE17573.

XX Composition comprising MTB39 antigen and MTB32A antigen from
 MYcobacterium species, useful for eliciting immune response in a subject.
 Claim 81; Page 108-109; 136pp; English.

XX The present invention relates to fusion proteins containing at least two
 MYcobacterium species antigens, nucleotides encoding them and
 compositions comprising such fusion proteins. The present invention
 particularly relates to nucleic acids encoding fusion proteins that
 include two or more individual M. tuberculosis antigens which increase
 the serological sensitivity of sera from individuals infected with
 tuberculosis and methods for their use in diagnosis, prevention and
 treatment of tuberculosis infection. Sequences of the invention are
 useful for eliciting an immune response in a mammal, e.g., human,
 immunised with BCG. They are useful in the diagnosis, treatment and
 prevention of Mycobacterium infection. The fusion proteins and the
 polynucleotides are useful as diagnostic tools in patients infected with
 Mycobacterium, in vitro and in vivo assays for detecting humoral
 antibodies or cell-mediated immunity against M. tuberculosis, for the
 diagnosis of an infection or monitoring of disease progression, as
 immunogens to generate or elicit a protective immune response in a
 patient and for raising anti-M. tuberculosis antibodies in a non-human
 animal. Sequences of the invention are also useful as vaccines. MTB32A
 fusion proteins of the invention are useful as in vivo diagnostic agents
 for intradermal skin test. The present sequence is a DNA encoding
 CC Mycobacterium species MTB72FmutSA (Ra12-TbH9-Ra35MutSA) mutant fusion
 CC protein.

XX Sequence 2190 BP; 340 A; 697 C; 797 G; 356 T; 0 U; 0 Other;

Query Match 95.5%; Score 2185.2; DB 6; Length 2190;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2187; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 42 ATGCATCACCATCACCATCACACGCGCGTCCGATTAACCTCCAGCTGCCAGGTTGGG 101

Db 1 ATGCATCACCATCACCATCACACGCGCGTCCGATTAACCTCCAGCTGCCAGGTTGGG 60

QY 102 CAGGATTTCGCATTCCTCCATCGGCGGCGATCGGATCGGCGGCGATCCGATCCGAT 161

Db 61 CAGGATTTCGCATTCCTCCATCGGCGGCGATCGGATCGGCGGCGATCCGATCCGAT 120

QY 162 GGGGGTCAACCCAGCTTCATATCGGCGCTACCGCTTCCTCGCTTGGGTGTTGTCGAC 221

Db 121 GGGGGTCAACCCAGCTTCATATCGGCGCTACCGCTTCCTCGCTTGGGTGTTGTCGAC 180

QY 222 AACACGCGACGCGGCGACGAGTCCAAACGGTGTGCGGAGCGCTCCGCGGCGAGTCTC 281

Db 181 AACACGCGACGCGGCGGCGACGAGTCCAAACGGTGTGCGGAGCGCTCCGCGGCGAGTCTC 240

QY 282 GGCATCTCCACCGCGGCGAGTGCATCACCGCGGTCCAGCGCGCTCCGATCAACTCGGCCACC 341

Db 241 GGCATCTCCACCGCGGCGAGTGCATCACCGCGGTCCAGCGCGCTCCGATCAACTCGGCCACC 300

QY 342 GCGATGGGCGGCGGCGCTTAACGGGCGATCATCCCGGTGACGTCATCTCGGTGACCTGGCAA 401

Db 301 GCGATGGCGGACGCGCTTAAAGGCGCATCATCCGGTGACGTCTCTCGGTGACCTGGCAA 360
Qy 402 ACCAAGTCGGGCGGACGCGCTTACAGGAAACGTGAATTTGGCGGAGGAGCCCGCGCGCAA 461
Db 361 ACCAAGTCGGGCGGACGCGCTTACAGGAAACGTGAATTTGGCGGAGGAGCCCGCGCGCAA 420
Qy 462 TTCATGGTCGATTTTCGGGCGGTTTACACCGGAGATCACTCCCGGAGGATGTACGCCGCG 521
Db 421 TTCATGGTCGATTTTCGGGCGGTTTACACCGGAGATCACTCCCGGAGGATGTACGCCGCG 480
Qy 522 CCGGGTTCCGGCCTCGTGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGGAGTGAACCTG 581
Db 481 CCGGGTTCCGGCCTCGTGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGGAGTGAACCTG 540
Qy 582 TTTTCGGCGCGTGGCGCTTTCAGTCGGTGGTCTGGGGTCTGACGGTGGGGTCTGGGATA 641
Db 541 TTTTCGGCGCGTGGCGCTTTCAGTCGGTGGTCTGGGGTCTGACGGTGGGGTCTGGGATA 600
Qy 642 GGTTCGTTCGGCGGCTTCGATGTGGCGCGGCTCGCCGTATGTGGCGTGGATGAGCGCTC 701
Db 601 GGTTCGTTCGGCGGCTTCGATGTGGCGCGGCTCGCCGTATGTGGCGTGGATGAGCGCTC 660
Qy 702 ACCCGGGGAGCGCGAGCTGACCGCGCGCCAGGTCCGGGTTGCTCGCGCGGCTTACGAG 761
Db 661 ACCCGGGGAGCGCGAGCTGACCGCGCGCCAGGTCCGGGTTGCTCGCGCGGCTTACGAG 720
Qy 762 ACCGGCTATGGGCTGACGCTGCGCGCGCGCTGATCGCGAGAACGCTGCTCAACTGATG 821
Db 721 ACCGGCTATGGGCTGACGCTGCGCGCGCGCTGATCGCGAGAACGCTGCTCAACTGATG 780
Qy 822 ATTCTGATAGCGACCAACCTCTTTGGGCGAAACACCCCGCGGATCGCGTCAACGAGGCC 881
Db 781 ATTCTGATAGCGACCAACCTCTTTGGGCGAAACACCCCGCGGATCGCGTCAACGAGGCC 840
Qy 882 GAATAGCGGAGATGTGGCGGCGGAGCGCGCGGATGTTGGCTACCGCGCGGAGCG 941
Db 841 GAATAGCGGAGATGTGGCGGCGGAGCGCGCGGATGTTGGCTACCGCGCGGAGCG 900
Qy 942 GCGACGCGGAGCGCGGCTTTCGCGGCTTCGAGGAGCGCGCGGAGATGACCGCGCGGT 1001
Db 901 GCGACGCGGAGCGCGGCTTTCGCGGCTTCGAGGAGCGCGCGGAGATGACCGCGCGGT 960
Qy 1002 GGGCTCTTCGAGGAGCGCGCGGCTTCGAGGAGCGCGCGGAGATGACCGCGCGGAGCG 1061
Db 961 GGGCTCTTCGAGGAGCGCGCGGCTTCGAGGAGCGCGCGGAGATGACCGCGCGGAGCG 1020
Qy 1062 TTGATGAACTATGTGCGCGGCGGCTGCAACAGCTGCGCGGAGCGCGCGGAGATGACCG 1121
Db 1021 TTGATGAACTATGTGCGCGGCGGCTGCAACAGCTGCGCGGAGCGCGCGGAGATGACCG 1080
Qy 1122 CCTTCTTCCAGCTGGTGGCTTGTGGAAGACGCTTCGCGGATCGGTGCGCGATCAGC 1181
Db 1081 CCTTCTTCCAGCTGGTGGCTTGTGGAAGACGCTTCGCGGATCGGTGCGCGATCAGC 1140
Qy 1182 AACATGTGTGATGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 1241
Db 1141 AACATGTGTGATGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 1200
Qy 1242 AACACCTTGAAGCTGAGTGTGGAAGGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCG 1301
Db 1201 AACACCTTGAAGCTGAGTGTGGAAGGCTTTCGCGGCGGCGGCGGCGGCGGCGGCGGCG 1260
Qy 1302 ACCCGCGGCGGAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 1361
Db 1261 ACCCGCGGCGGAGCGGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCGGAGCG 1320
Qy 1362 GGTCTGGCGGCTGGGCTGGCGGCGGAGCTTGGGTTCGGCGGCGGCTCGGTGCGGTTCG 1421
Db 1321 GGTCTGGCGGCTGGGCTGGCGGCGGAGCTTGGGTTCGGCGGCGGCTCGGTGCGGTTCG 1380
Qy 1422 GTGCCGAGCGCTGGCGCGGCGGCGGAGCTCAACGAGCTCAACCGCGCGGCGGCGGCGGCG 1481

Db 1381 GTGCCGAGCGCTGGCGCGGCGGCGGAGCTCAACCGGCGGCGGCGGCGGCGGCGGCGG 1440
Qy 1482 CTGACGAGCGCTGACCGCGCGGAGGAGGCGGCGGAGATGCTGGCGGCGGCGGCGGCGG 1541
Db 1441 CTGACGAGCGCTGACCGCGCGGAGGAGGCGGCGGAGATGCTGGCGGCGGCGGCGGCGG 1500
Qy 1542 GTGGGCGAGATGGCGCGGCGGCGGCGGCTCAGTGGTGTGCTGCTGCTGCTGCTGCTG 1601
Db 1501 GTGGGCGAGATGGCGCGGCGGCGGCGGCTCAGTGGTGTGCTGCTGCTGCTGCTGCTG 1560
Qy 1602 CGACCTATGTGATGCGGCGGCGGCGGCGGATATCGCCCGCGGCGGCGGCGGCGGCGG 1661
Db 1561 CGACCTATGTGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1620
Qy 1662 CAGGACCGGCTTCGCGGAGCTTCCCGCGGCTCGCCCTCGACCGCTCGCGGATGGTCCGCAA 1721
Db 1621 CAGGACCGGCTTCGCGGAGCTTCCCGCGGCTCGCCCTCGACCGCTCGCGGATGGTCCGCAA 1680
Qy 1722 GTGGGCGAGATGGCGCGGCGGCGGCGGCTCAGTGGTGTGCTGCTGCTGCTGCTGCTGCTG 1781
Db 1681 GTGGGCGAGATGGCGCGGCGGCGGCGGCTCAGTGGTGTGCTGCTGCTGCTGCTGCTGCTG 1740
Qy 1782 GGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1841
Db 1741 GGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1800
Qy 1842 GGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1901
Db 1801 GGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1860
Qy 1902 GTGGGCTATGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1961
Db 1861 GTGGGCTATGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1920
Qy 1962 CCGTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2021
Db 1921 CCGTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1980
Qy 2022 AGCGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2081
Db 1981 AGCGGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2040
Qy 2082 ACCGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2141
Db 2041 ACCGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2100
Qy 2142 TTCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2201
Db 2101 TTCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2160
Qy 2202 GTGGTGGGCTATGAAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2231
Db 2161 GTGGTGGGCTATGAAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2190

Search completed: July 3, 2004, 17:52:08

Job time : 910 secs

RESULT 2
US-09-287-849-1
; Sequence 1, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2287
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein Ral2-TbH3-Ral5 (designated Mb332-Mb39
; OTHER INFORMATION: fusion)
; NAME/KEY: modified_base
; LOCATION: (30)
; OTHER INFORMATION: n = g, a, c or t
; NAME/KEY: modified_base
; LOCATION: (33)
; OTHER INFORMATION: n = g, a, c or t
; NAME/KEY: CDS
; LOCATION: (42)..(2231)
; NAME/KEY: modified_base
; LOCATION: (2270)
; OTHER INFORMATION: n = g, a, c or t
US-09-287-849-1

Query Match 99.9%; Score 2284; DB 4; Length 2287;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2287; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	TCTAGAATAATTTTGTACTTTTAAAGANGANATATACATATGCATCACCATC	60
Db	1	TCTAGAATAATTTTGTACTTTTAAAGANGANATATACATATGCATCACCATC	60
Qy	61	ACACGGCGCGTCCGATAAATTCAGCTGTCCAGGGTGGGAGGATTCGCCATTCCGA	120
Db	61	ACACGGCGCGTCCGATAAATTCAGCTGTCCAGGGTGGGAGGATTCGCCATTCCGA	120
Qy	121	TCGGGCGGGGATGGGATTCGGGCGGAGATCCGATCGGTGGGGGTACCCACCGTTC	180
Db	121	TCGGGCGGGGATGGGATTCGGGCGGAGATCCGATCGGTGGGGGTACCCACCGTTC	180
Qy	181	ATATCGGCGCTTACCGCTTCTCGGCTTGGGTGTTCGACAAACAGCGAAGGGGCGAC	240
Db	181	ATATCGGCGCTTACCGCTTCTCGGCTTGGGTGTTCGACAAACAGCGAAGGGGCGAC	240
Qy	241	GAGTCCAAACGGCTCGGTGGGAGCGCTCCGGCGGCAAGTCTCGGATCTCCACCGGCGACG	300
Db	241	GAGTCCAAACGGCTCGGTGGGAGCGCTCCGGCGGCAAGTCTCGGATCTCCACCGGCGACG	300

Qy	301	TGATCACCGCGGTGACGCGGCTCCGATCAATCGCCACCGCGATGGCGGACGCGCTTA	360
Db	301	TGATCACCGCGGTGACGCGGCTCCGATCAATCGCCACCGCGATGGCGGACGCGCTTA	360
Qy	361	ACGGGATCATCCCGGTGACGCTCATCTCGTGACCTGGCAAAACCAAGTCGGGCGGACGC	420
Db	361	ACGGGATCATCCCGGTGACGCTCATCTCGTGACCTGGCAAAACCAAGTCGGGCGGACGC	420
Qy	421	GTACAGGGAACGTGACATTTGGCGAGGACCCCGCGCCGAATTCATGGTGGATTTGGGG	480
Db	421	GTACAGGGAACGTGACATTTGGCGAGGACCCCGCGCCGAATTCATGGTGGATTTGGGG	480
Qy	481	CGTTTACCACCGGAGATCAACTCCGCGAGGATGTACCGCGGCCCGGGTTCGGCTCGCTGG	540
Db	481	CGTTTACCACCGGAGATCAACTCCGCGAGGATGTACCGCGGCCCGGGTTCGGCTCGCTGG	540
Qy	541	TGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTGACCTGTTTTCGGCGCGCTCGCGCT	600
Db	541	TGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTGACCTGTTTTCGGCGCGCTCGCGCT	600
Qy	601	TTCACTCGGTGCTGGGGTCTGACCGTGGGGTCTGATAGTGTTCGCGCGGCTCTGA	660
Db	601	TTCACTCGGTGCTGGGGTCTGACCGTGGGGTCTGATAGTGTTCGCGCGGCTCTGA	660
Qy	661	TGTTGGGCGGCGCTCCCGTATGTGCGGTGAGTGCACCGCGGCGGCGGCGGCGGCG	720
Db	661	TGTTGGGCGGCGCTCCCGTATGTGCGGTGAGTGCACCGCGGCGGCGGCGGCGGCGG	720
Qy	721	TGACCGCGCGCGCTCGGCTTCTCGCGCGGCTACGAGACGCGCTATGGGTGTGACGG	780
Db	721	TGACCGCGCGCGCTCGGCTTCTCGCGCGGCTACGAGACGCGCTATGGGTGTGACGG	780
Qy	781	TGCCCCCGCGGTGATTCGCGAGAACCGTGTCTGAACTGATGATTCATAGCGACCAACC	840
Db	781	TGCCCCCGCGGTGATTCGCGAGAACCGTGTCTGAACTGATGATTCATAGCGACCAACC	840
Qy	841	TCCTGGGCGGCAACACACCGCGCGATTCGGGTCAACAGAGCGGCAATACGGCGAGATGGG	900
Db	841	TCCTGGGCGGCAACACACCGCGCGATTCGGGTCAACAGAGCGGCAATACGGCGAGATGGG	900
Qy	901	CCCAAGACGCGCGCGATTTTGGCTACGCGCGGCGACCGCGACGCGCGGCGGCGGCT	960
Db	901	CCCAAGACGCGCGCGATTTTGGCTACGCGCGGCGACCGCGACGCGCGGCGGCGGCT	960
Qy	961	TGCTCCGTTTCAGAGAGCGCGGAGATGACACGCGGGTGGGCTCTTCAGCAGGCGCG	1020
Db	961	TGCTCCGTTTCAGAGAGCGCGGAGATGACACGCGGGTGGGCTCTTCAGCAGGCGCG	1020
Qy	1021	CGCGGCTCAGAGAGCGCTCCGACACCGCGCGGCGAACCACTTGATGAACAATGTGCC	1080
Db	1021	CGCGGCTCAGAGAGCGCTCCGACACCGCGCGGCGAACCACTTGATGAACAATGTGCC	1080
Qy	1081	AGGCGCTGCAACAGCTGGCGCGACCGCGAGGCGCACCGCTTCTTCAAGCTGGGTG	1140
Db	1081	AGGCGCTGCAACAGCTGGCGCGACCGCGAGGCGCACCGCTTCTTCAAGCTGGGTG	1140
Qy	1141	GCTGTGGAAGACGCTTCGCGCGATCGGTCCCGATCAGCAACATGGTGTTCAGTGGCG	1200
Db	1141	GCTGTGGAAGACGCTTCGCGCGATCGGTCCCGATCAGCAACATGGTGTTCAGTGGCG	1200
Qy	1201	ACAACACATGTTCGATGACCAACTCGGCTGTTCATGACCAACACCTTGAGCTCGATG	1260
Db	1201	ACAACACATGTTCGATGACCAACTCGGCTGTTCATGACCAACACCTTGAGCTCGATG	1260
Qy	1261	TGAAGGGCTTTCCTCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1320
Db	1261	TGAAGGGCTTTCCTCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1320
Qy	1321	TCCGCGGATGAGCTCGGTGGGCGAGCTCGTGGGTTCCTCGGCTTGCGGCTTGCGG	1380
Db	1321	TCCGCGGATGAGCTCGGTGGGCGAGCTCGTGGGTTCCTCGGCTTGCGGCTTGCGG	1380

Db 564 GCTCTCGAGCAGCGCGCGGCTCGAGGAGCTCCGACACCGCGCGGAGAACAGTT 623
QY 1064 GATGAACAAATGTGCCCCAGCGCTGCAACAGCTGCGCCAGCCACCAACGAGGCAACAGCC 1123
Db 624 GATGAACAAATGTGCCCCAGCGGCTGCAACAGCTGCGCCAGCCACCAACGAGGCAACAGCC 683
QY 1124 TTTCTTCCAAAGTGGTGGCTGTGGAAGAGCGTCTCGCGCATCGGTCGCGGATCAGCAA 1183
Db 684 TTTCTTCCAAAGTGGTGGCTGTGGAAGAGCGTCTCGCGCATCGGTCGCGGATCAGCAA 743
QY 1184 CATGTGTGATGCGCAACCAACCAATGTGATGACCAATCGGTTGTGATGACCA 1243
Db 744 CATGTGTGATGCGCAACCAACCAATGTGATGACCAATCGGTTGTGATGACCA 803
QY 1244 CACCTTGTAGCTCGATGTTGAAGGCTTGTCTCGCGCGCGCGCCAGCGCGCTGCAAC 1303
Db 804 CACCTTGTAGCTCGATGTTGAAGGCTTGTCTCGCGCGCGCGCCAGCGCGCTGCAAC 863
QY 1304 CGCGCGCAAAACCGGGTTCGGGCGATGAGTCTGCTGGGAGTGTGCTGGGTTCTCGG 1363
Db 864 CGCGCGCAAAACCGGGTTCGGGCGATGAGTCTGCTGGGAGTGTGCTGGGTTCTCGG 923
QY 1364 TCTGGCGGCTGGGTCGCGCAACTTGGTTCGGGCGCGCTCGGTCGCTGCTGCTGCT 1423
Db 924 TCTGGCGGCTGGGTCGCGCAACTTGGTTCGGGCGCGCTCGGTCGCTGCTGCTGCT 983
QY 1424 GCGCAGGCTGGGTCGCGCAACTGAGGAGTCAACCGCGCGCGCGCGCGCTGCGCT 1483
Db 984 GCGCAGGCTGGGTCGCGCAACTGAGGAGTCAACCGCGCGCGCGCGCGCTGCGCT 1043
QY 1484 GACGAGCTGACCGCGCGGAGAGAGGCGCGCGAGTCTGCGCGCGCTGCGCT 1543
Db 1044 GACGAGCTGACCGCGCGGAGAGAGGCGCGCGAGTCTGCGCGCGCTGCGCT 1103
QY 1544 GGGCAGATGGCGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1603
Db 1104 GGGCAGATGGCGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1163
QY 1604 ACCCTATGATGCGGATCTCGGCGAGCGCGGATGCTGCGCGCGCTGCTGCTGCTGCT 1663
Db 1164 ACCCTATGATGCGGATCTCGGCGAGCGCGGATGCTGCGCGCGCTGCTGCTGCTGCT 1223
QY 1664 GAGCGGTCGCGGATCTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1723
Db 1224 GAGCGGTCGCGGATCTCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1283
QY 1724 GGGGCGAGGTCGATCAACCAACCACTGGGCTCAACCAACCGCGCGCGCGCGCG 1783
Db 1284 GGGGCGAGGTCGATCAACCAACCACTGGGCTCAACCAACCGCGCGCGCGCGCG 1343
QY 1784 GACGCGATGCTGATCCCAACCGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1843
Db 1344 GACGCGATGCTGATCCCAACCGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1403
QY 1844 GCGCAGCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1903
Db 1404 GCGCAGCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1463
QY 1904 GCGGATGACCGCGACCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1963
Db 1464 GCGGATGACCGCGACCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1523
QY 1964 GTCGCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2023
Db 1524 GTCGCGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1583
QY 2024 GGTGGGCGAGCGGAAACCGCCCTGCGGTCGCTGCGAGGCTGCTGCGCTCGGCGCAAC 2083
Db 1584 GGTGGGCGAGCGGAAACCGCCCTGCGGTCGCTGCGAGGCTGCTGCGCTCGGCGCAAC 1643
QY 2084 GGTGAGCGGTCGATTCGCTGAGCGGTCGCGAGAGACATGAAACGGGTTGATCAGTT 2143
Db 1644 GGTGAGCGGTCGATTCGCTGAGCGGTCGCGAGAGACATGAAACGGGTTGATCAGTT 1703

QY 2144 CGATGCGCGGATCCAGCCCGGCTGATTCGGCGCGCGCGCTCGTCAACGGCTAGGACAGGT 2203
Db 1704 CGATGCGCGGATCCAGCCCGGCTGATTCGGCGCGCGCGCTCGTCAACGGCTAGGACAGGT 1763
QY 2204 GGTGCGGATGAACACCGCGCGCTCCTAGGATATC 2237
Db 1764 GGTGCGGATGAACACCGCGCGCTCCTAGGATATC 1797

RESULT 4

US-09-287-849-21
; Sequence 21, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Read, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 1801
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein TbH9-DPv-MTI (designated Mt61f)
; NAME/KEY: CDS
; LOCATION: (1)..(1800)
US-09-287-849-21

Query Match: 51.3%; Score 1172.8; DB 4; Length 1801;
Best Local Similarity 99.0%; Pred. No. 6e-225;
Matches 1180; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 464 CATGCTGATTTCCGGCGGTTACACCGGAGATCACTCCGCGAGATGTACCGCGGCC 523
Db 24 CATGCTGATTTCCGGCGGTTACACCGGAGATCACTCCGCGAGATGTACCGCGGCC 83
QY 524 GGGTTCGCGCTCGTGTGTCGCGGCTCAGATGTGGGACAGCGTGGCGAGTACCTGTT 583
Db 84 GGGTTCGCGCTCGTGTGTCGCGGCTCAGATGTGGGACAGCGTGGCGAGTACCTGTT 143
QY 584 TTCGCGCGCTCGCGCTTCAGTCCGCTGCTGGGTCGCTGCGGTCGCGGTCGCGATAGG 643
Db 144 TTCGCGCGCTCGCGCTTCAGTCCGCTGCTGGGTCGCTGCGGTCGCGGTCGCGATAGG 203
QY 644 TTCGTCGCGGCTGATGTCGCGGCGGCTCGCGCTATGTCGCGTGTGATGAGCTCAC 703
Db 204 TTCGTCGCGGCTGATGTCGCGGCGGCTCGCGCTATGTCGCGTGTGATGAGCTCAC 263
QY 704 CGCGGGCAGCGGAGTCAACCGCGCGCGCGCGCGGTCGCGGTCGCGGTCGCGGTCGCGGAC 763
Db 264 CGCGGGCAGCGGAGTCAACCGCGCGCGCGCGGTCGCGGTCGCGGTCGCGGTCGCGGAC 323

764 GCGGTATGGGCTGACGGTCCCGCCGCGGTGATCGCGAGAACCGTGTGAACTGATGAT 823
Db
324 GCGGTATGGCTGACGGTCCCGCCGCGGTGATCGCGAGAACCGTGTGAACTGATGAT 383
Qy
824 TCTGATAGCGACCAACTCTTGGGGCAAAACCCCGCGGATCGCGGTCAACGAGCCGA 883
Db
384 TCTGATAGCGACCAACTCTTGGGGCAAAACCCCGCGGATCGCGGTCAACGAGCCGA 443
Qy
884 ATACGGCGAGATGTGGGCCCAAGACGCGCCGCGATGTTGGCTACGCGGGGACGGC 943
Db
444 ATACGGCGAGATGTGGGCCCAAGACGCGCCGCGATGTTGGCTACGCGGGGACGGC 503
Qy
944 GACGGCGAGCGGACGTTGCTCCGTTGAGAGAGCGCGGAGATGACACGCGGGTGG 1003
Db
504 GACGGCGAGCGGACGTTGCTCCGTTGAGAGAGCGCGGAGATGACACGCGGGTGG 563
Qy
1004 GTCTCTCGAGAGCGCGCGCGGTGAGAGAGCGCTCCGACACCGCGCGGCGAAGT 1063
Db
564 GTCTCTCGAGAGCGCGCGCGGTGAGAGAGCGCTCCGACACCGCGCGGCGAAGT 623
Qy
1064 GATGAACAATGTGCCCGCGGCTGCAACAGCTGGCCCGACCGCGAGGCGCACCGCC 1123
Db
624 GATGAACAATGTGCCCGCGGCTGCAACAGCTGGCCCGACCGCGAGGCGCACCGCC 683
Qy
1124 TTTCTTCCAGCTGGGTGGCTGTGGAAGACGGTCTCGCGCATCGGTGCGCGATCAGCAA 1183
Db
684 TTTCTTCCAGCTGGGTGGCTGTGGAAGACGGTCTCGCGCATCGGTGCGCGATCAGCAA 743
Qy
1184 CATGTGTGATGGGCAACCAACCATGTGATGACCAACTCGGTGTGTCGATGACCAA 1243
Db
744 CATGTGTGATGGGCAACCAACCATGTGATGACCAACTCGGTGTGTCGATGACCAA 803
Qy
1244 CACCTTGAGCTCGATGTTGAAGGGCTTGTCTCGGGCGCGCCCGCGAGCGCGTCAAAC 1303
Db
804 CACCTTGAGCTCGATGTTGAAGGGCTTGTCTCGGGCGCGCCCGCGAGCGCGTCAAAC 863
Qy
1304 CGCGCGCAAAACCGGGTCCCGGCGATGAGCTCGCTGGGAGCTCGCTGGTCTTTCGGG 1363
Db
864 CGCGCGCAAAACCGGGTCCCGGCGATGAGCTCGCTGGGAGCTCGCTGGTCTTTCGGG 923
Qy
1364 TCTGGGCGGTGGGTGGCGGCAACTTGGGTGGGCGGCGCTCGGTGCGTTCGTTGCGT 1423
Db
924 TCTGGGCGGTGGGTGGCGGCAACTTGGGTGGGCGGCGCTCGGTGCGTTCGTTGCGT 983
Qy
1424 GCCGAGCGCTGGCGCGGCGCAACCGAGTACACCGCGCGCGCGCGCGCTGCGCT 1483
Db
984 GCCGAGCGCTGGCGCGGCGCAACCGAGTACACCGCGCGCGCGCGCGCTGCGCT 1043
Qy
1484 GACGAGCTGACAGCGCGCGGCAACCGAGTACACCGCGCGCGCGCGCGCTGCGCT 1543
Db
1044 GACGAGCTGACAGCGCGCGGCAACCGAGTACACCGCGCGCGCGCGCGCTGCGCT 1103
Qy
1544 GGGGAGATGGGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1603
Db
1104 GGGGAGATGGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1163
Qy
1604 ACCCTATGATGCGGATTTCTCGGCGAGCGCGGATATCGCCCGCGCGCGCGCGCG 1655
Db
1164 ACCCTATGATGCGGATTTCTCGGCGAGCGCGGATGATCCCGTGAGC 1215

RESULT 5
US-08-818-112-106
; Sequence 106, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 106:
SEQUENCE CHARACTERISTICS:
LENGTH: 3058 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-106

Query Match 51.1%; Score 1169; DB 3; Length 3058;
Best Local Similarity 99.6%; Pred. No. 3.7e-224;
Matches 11/2; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
Qy 463 TCATGTTGATTTTCGGGCGGTATCCACCGAGATCACTCCCGAGGATGACCGCGGCC 522
Db 391 TAATGTTGATTTTCGGGCGGTATCCACCGAGATCACTCCCGAGGATGACCGCGGCC 450
Qy 523 CGGGTTCGGCTCGTGTGGCGCGCTCAGATGTGGGACAGCGTGGCGAGTGAACCTGT 582
Db 451 CGGGTTCGGCTCGTGTGGCGCGCTCAGATGTGGGACAGCGTGGCGAGTGAACCTGT 510
Qy 583 TTTCCGCGCGCTCGGCGGTTTCAGTTCGGTGTCTCGGGTCTCAGTGGGTCGTTGGATG 642
Db 511 TTTCCGCGCGCTCGGCGGTTTCAGTTCGGTGTCTCGGGTCTCAGTGGGTCGTTGGATG 570
Qy 643 GTTCGTCGGCGGCTCGATGTGGCGCGGCTCGCGCTATGTGGCTGTGATGAGCGTCA 702
Db 571 GTTCGTCGGCGGCTCGATGTGGCGCGGCTCGCGCTATGTGGCTGTGATGAGCGTCA 630
Qy 703 CGGGGGGACGCGAGCTGACCGCGCGCGGCTCGCGCTATGTGGCTGTGATGAGCGTCA 762
Db 631 CGGGGGGACGCGAGCTGACCGCGCGCGGCTCGCGCTATGTGGCTGTGATGAGCGTCA 690
Qy 763 CGGCGTATGGGCTGACCGTGCCTCGCGCGGCTCGCGCTATGTGGCTGTGATGAGCGTCA 822
Db 691 CGGCGTATGGGCTGACCGTGCCTCGCGCGGCTCGCGCTATGTGGCTGTGATGAGCGTCA 750
Qy 823 TTTCTGATAGCGACCAACTCTTGGGGCAAAACACCGCGCGGATGCGCGTCAACGAGCGG 882
Db 751 TTTCTGATAGCGACCAACTCTTGGGGCAAAACACCGCGCGGATGCGCGTCAACGAGCGG 810
Qy 883 AATACGGCGAGATGTGGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 942
Db 811 AATACGGCGAGATGTGGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 870
Qy 943 CGAGCGGCGAGCGGCGGCTTCTGCTGCTTTCAGGAGGCGCGCGGATGACGAGCGGGTG 1002
Db 871 CGAGCGGCGAGCGGCGGCTTCTGCTGCTTTCAGGAGGCGCGCGGATGACGAGCGGGTG 930

1003 GGTCTCTCAGCAGCGCGCGGTGTCAGAGGCTCTCCGACACCGCGCGGGAACCAAGT 1062
Db GGTCTCTCAGCAGCGCGCGGTGTCAGAGGCTCTCCGACACCGCGCGGGAACCAAGT 990
1063 TGATGAACAATGTGCCAGGCGGTGCAACAGTGGCCAGCCGACGAGGCGACCAAGC 1122
Db TGATGAACAATGTGCCAGGCGGTGCAACAGTGGCCAGCCGACGAGGCGACCAAGC 1050
1123 CTTCTTCCAGCTGGGTGGCGGTGGAAGACGCTCTCGCGCATGGTGGCGGATCAGCA 1182
Db CTTCTTCCAGCTGGGTGGCGGTGGAAGACGCTCTCGCGCATGGTGGCGGATCAGCA 1110
1183 ACATGCTGCTGATGCGCAACCAACATGTCATGACCAACTCGGCTGGTGTGATGACA 1242
Db ACATGCTGCTGATGCGCAACCAACATGTCATGACCAACTCGGCTGGTGTGATGACA 1170
1243 ACACCTTGAGCTGATGTTGAAGGCTTTGCTCGCGCGCGCGCGCGCGCGCGGCAAA 1302
Db ACACCTTGAGCTGATGTTGAAGGCTTTGCTCGCGCGCGCGCGCGCGCGCGCGGCAAA 1230
1303 CCGCGCGGCAAAACCGGCGTCCGCGCATGAGCTCGCTGGGCGAGCTCGCTGGGTTCTTCGG 1362
Db CCGCGCGGCAAAACCGGCGTCCGCGCATGAGCTCGCTGGGCGAGCTCGCTGGGTTCTTCGG 1290
1363 GTCTGGCGGTGGGTGGCGCGCAACTTGGGTGGCGCGCGCTCGCTGGGTGGGTGGGTGG 1422
Db GTCTGGCGGTGGGTGGCGCGCAACTTGGGTGGCGCGCGCTCGCTGGGTGGGTGGGTGG 1350
1423 TGCGCGAGGCTGGCGCGCGCGCAACAGCAGTCAACCGCGCGCGCGCGCGCGCGCGCG 1482
Db TGCGCGAGGCTGGCGCGCGCGCAACAGCAGTCAACCGCGCGCGCGCGCGCGCGCGCG 1410
1483 TGACCAAGCTTACAGCGCGCGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1542
Db TGACCAAGCTTACAGCGCGCGCGCAAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1470
1543 TGGGCGAGTGGCG 1602
Db TGGGCGAGTGGCG 1530
1603 GACCTATGTGATGCGCGCATTCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1639
Db GACCTATGTGATGCGCGCATTCCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1567

RESULT 6
US-08-818-111-101
; Sequence 101, Application US/08818111
; Patent No. 633852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 101:
SEQUENCE CHARACTERISTICS:
LENGTH: 3058 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-101

Query March 51.1%; Score 1169; DB 4; Length 3058;
Best Local Similarity 99.6%; Pred. No. 3.7e-224;
Matches 1172; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 463 TCATGTTGATTTTCGGGCGTTTACACCGGAGATCAACTCCGCGAGGATGTACCGCGCC 522
Db TAATGTTGATTTTCGGGCGTTTACACCGGAGATCAACTCCGCGAGGATGTACCGCGCC 450
QY 523 CGGTTTCGGGCTTCGCTGGTGGCGCGCTCAGATGTGGAGCAGCGTGGCGAGTCACTGT 582
Db CGGTTTCGGGCTTCGCTGGTGGCGCGCTCAGATGTGGAGCAGCGTGGCGAGTCACTGT 510
QY 583 TTTTCGGCGCGCTTCGCTGGTGGTTCAGTTCGTTCTGGGCTCTGACCGTGGGTCGTGGATG 642
Db TTTTCGGCGCGCTTCGCTGGTGGTTCAGTTCGTTCTGGGCTCTGACCGTGGGTCGTGGATG 570
QY 643 GTTCGTCGGCGGCTTCGATGTGGGCGCGCTCGCGGTATGTGGCTGTGGATGAGCGTCA 702
Db GTTCGTCGGCGGCTTCGATGTGGGCGCGCTCGCGGTATGTGGCTGTGGATGAGCGTCA 630
QY 703 CGCGGCGCAGCGCGAGTGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 762
Db CGCGGCGCAGCGCGAGTGTACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 690
QY 763 CGCGGTATGGGTGACCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 822
Db CGCGGTATGGGTGACCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 750
QY 823 TTCTGATAGCGACCAACCTCTTGGGCGCAACACCGCGCGCGCGCGCGCGCGCGCGCG 882
Db TTCTGATAGCGACCAACCTCTTGGGCGCAACACCGCGCGCGCGCGCGCGCGCGCGCG 810
QY 883 AATACGCGAGATGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 942
Db AATACGCGAGATGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 870
QY 943 CGACGCGCAGCGCGAGCTTGTGCGGTTTCGAGGAGCGCGCGCGCGCGCGCGCGCGCG 1002
Db CGACGCGCAGCGCGAGCTTGTGCGGTTTCGAGGAGCGCGCGCGCGCGCGCGCGCGCG 930
QY 1003 GGTCTCTCAGCAGCG 1062
Db GGTCTCTCAGCAGCG 990
QY 1063 TGATGAACAATGTGCCAGGCGGTGCAACAGTGGCCAGCCGACGAGGCGCGCGCGCG 1122
Db TGATGAACAATGTGCCAGGCGGTGCAACAGTGGCCAGCCGACGAGGCGCGCGCGCG 1050
QY 1123 CTTCTTCCAGCTGGGTGGCGGTGGAAGACGCTCTCGCGCATGGTGGCGGATCAGCA 1182
Db CTTCTTCCAGCTGGGTGGCGGTGGAAGACGCTCTCGCGCATGGTGGCGGATCAGCA 1110
QY 1183 ACATGCTGCTGATGCGCAACCAACATGTCATGACCAACTCGGCTGGTGTGATGACA 1242
Db ACATGCTGCTGATGCGCAACCAACATGTCATGACCAACTCGGCTGGTGTGATGACA 1170

QY 1243 ACACCTTGAGCTCGATGTTGAAGGGCTTGTCTCCGGCGGCGCGCCCGCAGGCGCTGCAAA 1302
DB 1171 ACACCTTGAGCTCGATGTTGAAGGGCTTGTCTCCGGCGGCGCGCCCGCAGGCGCTGCAAA 1230
QY 1303 CGCGCGCAAAACCGGGTCCGGGCGATGAGTCTCGCTGGGCGAGCTCGCTGGGCTTCTCGG 1362
DB 1231 CGCGCGCAAAACCGGGTCCGGGCGATGAGTCTCGCTGGGCGAGCTCGCTGGGCTTCTCGG 1290
QY 1363 GTCTGGGCGCTGGGTCGGCGCGCAACTTGGGTCGGGCGGCGCTCGCTGGGCTTCTCGG 1422
DB 1291 GTCTGGGCGCTGGGTCGGCGCGCAACTTGGGTCGGGCGGCGCTCGCTGGGCTTCTCGG 1350
QY 1423 TGCCGAGGCTGGGCGCGCGCGCAACACGAGGAGTCAACCCCGCGCGCGCGCGCTGCGCG 1482
DB 1351 TGCCGAGGCTGGGCGCGCGCGCAACACGAGGAGTCAACCCCGCGCGCGCGCGCTGCGCG 1410
QY 1483 TGACGAGCTGACGAGCGCGCGGGAAGAGGCGCGGCGAGATGCTGGGCGGCGCTGCGG 1542
DB 1411 TGACGAGCTGACGAGCGCGCGGGAAGAGGCGCGGCGAGATGCTGGGCGGCGCTGCGG 1470
QY 1543 TGGGCGAGATGGGCGCGCGCGCGTGGTGGGCTGAGTGGTGGTGGTGGTGGTGGTGGTGG 1602
DB 1471 TGGGCGAGATGGGCGCGCGCGCGTGGTGGGCTGAGTGGTGGTGGTGGTGGTGGTGGTGG 1530
QY 1603 GACCCATGATGCGCGCAATTCGCGGAGCGCGGA 1639
DB 1531 GACCCATGATGCGCGCAATTCGCGGCGCGCGGTA 1567

RESULT 7

US-09-556-106
; Sequence 106, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.M.
; APPLICANT: Dillon, David C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; TREATMENT
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3058 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-556-106

Query Match 51.1%; Score 1169; DB 4; Length 3058;
Best Local Similarity 99.6%; Pred. No. 3.7e-224;

Matches 1172; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 463 TCATGCTGATTTCCGGGCGTTTACACCGAGATCAACTCCCGGAGGATGTACGCGGCC 522
DB 391 TAATGCTGATTTCCGGGCGTTTACACCGAGATCAACTCCCGGAGGATGTACGCGGCC 450
QY 523 CGGCTTCCGCTCGCTGGTGGCGCGCGCTCAGATGTGGGACAGCGCTGGGAGTCACTGT 582
DB 451 CGGCTTCCGCTCGCTGGTGGCGCGCGCTCAGATGTGGGACAGCGCTGGGAGTCACTGT 510
QY 583 TTTCGCGCGCTGGGCTTTCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 642
DB 511 TTTCGCGCGCTGGGCTTTCAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 570
QY 643 GTTCGCTGGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 702
DB 571 GTTCGCTGGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 630
QY 703 CGCGGGGCGAGCGAGCTGACCGCGCGCGAGTCCGGTTCGCGCGCGCTTACGAGA 762
DB 631 CGCGGGGCGAGCGAGCTGACCGCGCGCGAGTCCGGTTCGCGCGCGCTTACGAGA 690
QY 763 CGCGCTGATGGGCTGACCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 822
DB 691 CGCGCTGATGGGCTGACCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 750
QY 823 TTTCGATAGCGACCAACCTTCTTGGGGCAAAACACCGCGCGATCGCGTCAACGAGCGG 882
DB 751 TTTCGATAGCGACCAACCTTCTTGGGGCAAAACACCGCGCGATCGCGTCAACGAGCGG 810
QY 883 AATAGCGAGATGTGGGCGCAAGACCGCGCGCGATGTTGGCTACCGCGGCGCGCGG 942
DB 811 AATAGCGAGATGTGGGCGCAAGACCGCGCGCGATGTTGGCTACCGCGGCGCGCGG 870
QY 943 CGACGCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1002
DB 871 CGACGCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 930
QY 1003 GCTCTCTGAGCAGCGCGCGGCTGCGAGGAGCTTCCGACACCGCGCGCGCGCGCGG 1062
DB 931 GCTCTCTGAGCAGCGCGCGGCTGCGAGGAGCTTCCGACACCGCGCGCGCGCGCGG 990
QY 1063 TGATGAACAATGTGCCCGCAGCGCTGCAACAGCTGGCGCGCGCGCGCGCGCGCGG 1122
DB 991 TGATGAACAATGTGCCCGCAGCGCTGCAACAGCTGGCGCGCGCGCGCGCGCGCGG 1050
QY 1123 CTCTTCCAAAGCTGGGCTGCTGGAAGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1182
DB 1051 CTCTTCCAAAGCTGGGCTGCTGGAAGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1110
QY 1183 ACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1242
DB 1111 ACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1170
QY 1243 ACACCTTGAGCTCGATGTTGAAGGGCTTGTCTCCGGGCGGCGCGCGCGCGCGGCA 1302
DB 1171 ACACCTTGAGCTCGATGTTGAAGGGCTTGTCTCCGGGCGGCGCGCGCGCGCGGCA 1230
QY 1303 CGCGGCGCAAAACCGGGTCCGGGCGATGAGTCTGCTGGGCGAGTCTGCTGGGCTTCTCGG 1362
DB 1231 CGCGGCGCAAAACCGGGTCCGGGCGATGAGTCTGCTGGGCGAGTCTGCTGGGCTTCTCGG 1290
QY 1363 GTCTGGGCGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1422
DB 1291 GTCTGGGCGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGGGCTGG 1350
QY 1423 TGCGCGAGGCTGGGCGCGCGCGCAACAGGAGTCAACCGCGCGCGCGCGCGCGCGCGG 1482
DB 1351 TGCGCGAGGCTGGGCGCGCGCGCAACAGGAGTCAACCGCGCGCGCGCGCGCGCGCGG 1410
QY 1483 TGACGAGCTGACGAGCGCGCGGGAAGAGGCGCGCGCGAGATGCTGGGCGGCGTGGCGG 1542
DB 1411 TGACGAGCTGACGAGCGCGCGGGAAGAGGCGCGCGCGAGATGCTGGGCGGCGTGGCGG 1470

GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Lodes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072,967
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 106:
SEQUENCE CHARACTERISTICS:
LENGTH: 3058 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

US-09-072-967-106

Query Match 51.1%; Score 1169; DB 4; Length 3058;
Best Local Similarity 99.6%; Pred. No. 3,7e-224;
Matches 1172; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 463 TCATGTTGGATTTCGGGGCGTTACACGGAGATCAACTCGGAGGATGACCGCGCC 522
Db 391 TAATGTTGGATTTCGGGGCGTTACACGGAGATCAACTCGGAGGATGACCGCGCC 450
QY 523 CGGGTTCCGGCTCGCTGGTGGCCCGCTCAGATGTGGGACAGCGTGGCGAGTGACTGT 582
Db 451 CGGGTTCCGGCTCGCTGGTGGCCCGCTCAGATGTGGGACAGCGTGGCGAGTGACTGT 510
QY 583 TTTCGGCGCGTCCGGCTTCAGTCGCTGGTGTGGGTCTGACGGTGGGTCTGTGATAG 642
Db 511 TTTCGGCGCGTCCGGCTTCAGTCGCTGGTGTGGGTCTGACGGTGGGTCTGTGATAG 570
QY 643 GTTCGTGGCGGCTCATGTGGTGGCGGCTCGCGGTATGTGGGTGATGAGGTCA 702
Db 571 GTTCGTGGCGGCTCATGTGGTGGCGGCTCGCGGTATGTGGGTGATGAGGTCA 630
QY 703 CCGCGGCGAGCGAGCTGACCGCGCCAGGTTCGGGTGTGGGTGGCGGCTTACGAGA 762
Db 631 CCGCGGCGAGCGAGCTGACCGCGCCAGGTTCGGGTGTGGGTGGCGGCTTACGAGA 690
QY 763 CGGGGTATGGCTACCGGTGCGCGCGGTGTGCGGAGAACCGTGTGACTGATGA 822
Db 691 CGGGGTATGGCTACCGGTGCGCGCGGTGTGCGGAGAACCGTGTGACTGATGA 750
QY 823 TTCTGATAGCAACCACTCTTGGGGCAAAACACCCCGCGATCGCGGTCAACGAGCGCG 882

Db 751 TTCTGATAGCAACCACTCTTGGGGCAAAACACCCCGCGATCGCGTCAACGAGCGCG 810
QY 883 AATAGCGGAGATGTGGGCGCCCAAGACCGCGCGCGATGTTGGCTTACCCCGCGCGCGG 942
Db 811 AATAGCGGAGATGTGGGCGCCCAAGACCGCGCGCGATGTTGGCTTACCCCGCGCGCGG 870
QY 943 CGACGCGACCGCGAGCTGTGCTGCGGTTGAGGAGCGCGCGAGATGACGAGCGCGGTG 1002
Db 871 CGACGCGACCGCGAGCTGTGCTGCGGTTGAGGAGCGCGCGAGATGACGAGCGCGGTG 930
QY 1003 GGCTCTCGAGCAGCGCGCGGTCGAGGAGCGCTCCGACACCGCGCGCGCAACAGT 1062
Db 931 GGCTCTCGAGCAGCGCGCGGTCGAGGAGCGCTCCGACACCGCGCGCGCAACAGT 990
QY 1063 TGATGAACAATGTGCCCGAGCGCTGCAACAGCTGGCGCCACCGCCACGAGGGCACCGC 1122
Db 991 TGATGAACAATGTGCCCGAGCGCTGCAACAGCTGGCGCCACCGCCACGAGGGCACCGC 1050
QY 1123 CTTCTTCCAAAGCTGGTGGCTGTGGAAGAGCGTCTCGCGCATCGGTCCGCGATCAGCA 1182
Db 1051 CTTCTTCCAAAGCTGGTGGCTGTGGAAGAGCGTCTCGCGCATCGGTCCGCGATCAGCA 1110
QY 1183 ACATGTTGTCGATGGCCCAACCAACACATGTGATGACCAACTCGGGTGTGTCGATGACCA 1242
Db 1111 ACATGTTGTCGATGGCCCAACCAACACATGTGATGACCAACTCGGGTGTGTCGATGACCA 1170
QY 1243 ACACCTTGAGCTCGATGTGAAGGCTTTGCTCGCGCGCGCGCCCGCGCGCGTGCAAA 1302
Db 1171 ACACCTTGAGCTCGATGTGAAGGCTTTGCTCGCGCGCGCGCCCGCGCGCGTGCAAA 1230
QY 1303 CGCGCGCGCAAAACCGGGTCCGGGCGATGAGTCTGGGCGAGCTCGGTGGGTCTTTCGG 1362
Db 1231 CGCGCGCGCAAAACCGGGTCCGGGCGATGAGTCTGGGCGAGCTCGGTGGGTCTTTCGG 1290
QY 1363 GTCTGGGCGGTGGGTGGCGCGCAACTTGGGTGGGCGGCTCGGTGGGTGTGTCGG 1422
Db 1291 GTCTGGGCGGTGGGTGGCGCGCAACTTGGGTGGGCGGCTCGGTGGGTGTGTCGG 1350
QY 1423 TGCGCGCAGCGCTGGGCGCGCGCAACAGGAGTCAACCGCGCGCGCGCGCGCGTGC 1482
Db 1351 TGCGCGCAGCGCTGGGCGCGCGCAACAGGAGTCAACCGCGCGCGCGCGCGCGTGC 1410
QY 1483 TGACCGCTTGACAGCGCGCGCGCAACAGGAGCGCGCGCGCGCGCGTGC 1542
Db 1411 TGACCGCTTGACAGCGCGCGCGCAACAGGAGCGCGCGCGCGCGCGTGC 1470
QY 1543 TGGGCGAGATGGCGCGCGCGCGCGGTGGTGGGTGAGTGGTGGTGGTGGTGGTGGTGG 1602
Db 1471 TGGGCGAGATGGCGCGCGCGCGCGGTGGTGGGTGAGTGGTGGTGGTGGTGGTGGTGG 1530
QY 1603 GACCCATGTGATGCGCGATCTCCGGCAGCGCGCGA 1639
Db 1531 GACCCATGTGATGCGCGATCTCCGGCAGCGCGCGA 1567

RESULT 10
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2


```
/ LENGTH: 4403765
/ TYPE: DNA
/ ORGANISM: Mycobacterium tuberculosis
/ FEATURE:
/ OTHER INFORMATION: CDC 1551
/ OTHER INFORMATION: "n" bases at various positions throughout the sequence
/ OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 51.1%; Score 1169; DB 3; Length 4403765;
Best Local Similarity 99.6%; Pred. No. 1.2e-223;
Matches 1172; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 463 TCATGTGTGATTTCCGGGGCTTACCAACCGAGATCAACTCCGAGAGATGTACGCCGCC 522
Db 1338813 TAATGTGTGATTTCCGGGGCTTACCAACCGAGATCAACTCCGAGAGATGTACGCCGCC 1338872
QY 523 CGGGTTCCGCTCGCTGTTGGCGCGCTCAGATGTGGACAGCGTGGCGAGTGAACCTGT 582
Db 1338873 CGGGTTCCGCTCGCTGTTGGCGCGCTCAGATGTGGACAGCGTGGCGAGTGAACCTGT 1338932
QY 583 TTTCGCGCGGCTCGGCTTTCAGTCGCTGTTCTGGGGTCTGACGCTGGGGTCTGTGATAG 642
Db 1338933 TTTCGCGCGGCTCGGCTTTCAGTCGCTGTTCTGGGGTCTGACGCTGGGGTCTGTGATAG 1338992
QY 643 GTTCGTTCCGCTCGCTGTTGGCGCGCTCAGATGTGGACAGCGTGGCGAGTGAACCTGT 702
Db 1338993 GTTCGTTCCGCTCGCTGTTGGCGCGCTCAGATGTGGACAGCGTGGCGAGTGAACCTGT 1339052
QY 703 CCGCCGGGAGCGCGAGCTGACCGCGCGCCAGCTCCGGGTTGCTCGCGCGGCTTACGAGA 762
Db 1339053 CCGCCGGGAGCGCGAGCTGACCGCGCGCCAGCTCCGGGTTGCTCGCGCGGCTTACGAGA 1339112
QY 763 CGCGCTATGGCTGACGCTGTCGCGCGCGCGCGCTGATCCCGAGAACCGTGTGAACGTGA 822
Db 1339113 CGCGCTATGGCTGACGCTGTCGCGCGCGCGCGCTGATCCCGAGAACCGTGTGAACGTGA 1339172
QY 823 TTCTGATAGGACCAACTCTTGGGGCAAAAACACCCCGCGATCGCGGTCAACGAGCGCG 882
Db 1339173 TTCTGATAGGACCAACTCTTGGGGCAAAAACACCCCGCGATCGCGGTCAACGAGCGCG 1339232
QY 883 AATACGGCGAGATGTGGGCGCAAGACCGCGCGCGATGTTTGGCTACGCGCGGCGAGCG 942
Db 1339233 AATACGGCGAGATGTGGGCGCAAGACCGCGCGCGATGTTTGGCTACGCGCGGCGAGCG 1339292
QY 943 CGACGGCGAGCGGAGCTGCTGCGCTTCGAGAGCGCGCGGAGATGACCGCGGGTG 1002
Db 1339293 CGACGGCGAGCGGAGCTGCTGCGCTTCGAGAGCGCGCGGAGATGACCGCGGGTG 1339352
QY 1003 GGCTCCTCGAGCAGCGCGCGCGCTCGAGAGCGCTCCGACACCGCGCGCGGAGCAACCA 1062
Db 1339353 GGCTCCTCGAGCAGCGCGCGCGCTCGAGAGCGCTCCGACACCGCGCGGCGAGCAACCA 1339412
QY 1063 TGATGAACATGTGCGCGAGCGCTGCAACAGCTGCGCGCGCGCGCGAGCGAGCGAGCG 1122
Db 1339413 TGATGAACATGTGCGCGAGCGCTGCAACAGCTGCGCGCGCGCGCGAGCGAGCGAGCG 1339472
QY 1123 CTTCCTTCAAGCTGGGTGGCTGTGAAGAAGCGCTCTCGCGGATCGGTCCCGGATCAGCA 1182
Db 1339473 CTTCCTTCAAGCTGGGTGGCTGTGAAGAAGCGCTCTCGCGGATCGGTCCCGGATCAGCA 1339532
QY 1183 ACATGTGTGATGCGCAACCAACATGTCGATGACCAACTCGGGTGTGTGATGACCA 1242
Db 1339533 ACATGTGTGATGCGCAACCAACATGTCGATGACCAACTCGGGTGTGTGATGACCA 1339592
QY 1243 ACACCTTGAGCTGATGTTGAGGGCTTTCCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1302
Db 1339593 ACACCTTGAGCTGATGTTGAGGGCTTTCCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1339652
QY 1303 CCGCCGCGCAAAAACGGGGTCCGGCGAGTGAAGCTGCTGGGCGAGCTCGCTGGGTCTTTCGG 1362
Db 1339653 CCGCCGCGCAAAAACGGGGTCCGGCGAGTGAAGCTGCTGGGCGAGCTCGCTGGGTCTTTCGG 1339712
```

Db 1339705 TTCTGATAGCACCACCTCTTGGGGCAAAACACCCCGCGATCGCGTCAACGAGCGCG 1339764
QY 883 AATACGGCGAGATGGGCCCAACAGCCCGCGGATGTTTGGCTACGCCGCGCGGACCG 942
Db 1339765 AATACGGCGAGATGGGCCCAACAGCCCGCGGATGTTTGGCTACGCCGCGCGGACCG 1339824
QY 943 CGAGCGGCGACCGGACGCTTGTGCGGTTTGGAGAGGCGCGCGAGATGACACGCGCGGCTG 1002
Db 1339825 CGAGCGGCGACCGGACGCTTGTGCGGTTTGGAGAGGCGCGCGAGATGACACGCGCGGCTG 1339884
QY 1003 GGCTCTTCGAGCAGCGCGCGCGGTCGAGAGGCGCTCGACACCGCGCGCGGACGACG 1062
Db 1339885 GGCTCTTCGAGCAGCGCGCGCGGTCGAGAGGCGCTCGACACCGCGCGCGGACGACG 1339944
QY 1063 TGATGAACAAATGTCGCCAGCGCTGCAACAGCTGGGCCAGCCACGACGCGGACCAACGC 1122
Db 1339945 TGATGAACAAATGTCGCCAGCGCTGCAACAGCTGGGCCAGCCACGACGCGGACCAACGC 1340004
QY 1123 CTTCTTCCAAAGCTGGGTGGCTGTGGAAGACGCTTCTCGCGCATCGGTTCGCCGATCAGCA 1182
Db 1340005 CTTCTTCCAAAGCTGGGTGGCTGTGGAAGACGCTTCTCGCGCATCGGTTCGCCGATCAGCA 1340064
QY 1183 ACATGGTCTCGATCGCAACCAACACATGTCGATGACCACTCGGCTGGTTCGATGACCA 1242
Db 1340065 ACATGGTCTCGATCGCAACCAACACATGTCGATGACCACTCGGCTGGTTCGATGACCA 1340124
QY 1243 ACACCTTTGAGCTCGATGTTGAAGGCTTTGTCGCGCGCGCGCGCGCGCGCTGCAAA 1302
Db 1340125 ACACCTTTGAGCTCGATGTTGAAGGCTTTGTCGCGCGCGCGCGCGCGCGCTGCAAA 1340184
QY 1303 CCGCGGCGCAAAACGGGTTCGGGCGATGAGCTGCTGGCAGCTGCTGGGTCTTCG 1362
Db 1340185 CCGCGGCGCAAAACGGGTTCGGGCGATGAGCTGCTGGGCGAGCTGCTGGGTCTTCG 1340244
QY 1363 GTCTGGGCGGTGGGTGGCGCCAACTTTGGGTTCGGGCGCGCTCGGTTCGTTTCG 1422
Db 1340245 GTCTGGGCGGTGGGTGGCGCCAACTTTGGGTTCGGGCGCGCTCGGTTCGTTTCG 1340304
QY 1423 TGCGCGAGGCTGGGCGCGCGCGCAACAGAGCAGTCAACCGCGCGCGCGCGCGCTGCGC 1482
Db 1340305 TGCGCGAGGCTGGGCGCGCGCGCAACAGAGCAGTCAACCGCGCGCGCGCGCGCTGCGC 1340364
QY 1483 TGACAGGCTGACAGCGCGCGCAAGAGAGGCGCGCGCGAGATGCTGGGCGGCTGCGCG 1542
Db 1340365 TGACAGGCTGACAGCGCGCGCAAGAGAGGCGCGCGCGAGATGCTGGGCGGCTGCGCG 1340424
QY 1543 TGGGCGAGATGGCG 1602
Db 1340425 TGGGCGAGATGGCG 1340484
QY 1603 GACCTATGATGCGCGCTTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1639
Db 1340485 GACCTATGATGCGCGCTTCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1340521

RESULT 12

US-09-103-840A-2/c
; Sequence 2, Application US/091039840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2

LENGTH: 4403765
TYPE: DNA
ORGANISM: Mycobacterium tuberculosis
FEATURE:
OTHER INFORMATION: CDC 1551
OTHER INFORMATION: "n" bases at various positions throughout the sequence
OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Query Match 41.4%; Score 947.4; DB 3; Length 4403765;
Best Local Similarity 89.3%; Pred. No. 1.7e-179;
Matches 1062; Conservative 0; Mismatches 106; Indels 21; Gaps 3;

QY 466 TGTGCGATTTCCGGGCGTTTACACCGGAGATCAACTCCGCGAGGATGTACCCGCGCGCGG 525
Db 1534826 TGTGCGATTTCCGGGCGTTTACACCGGAGATCAACTCCGCGAGGATGTATGCCGCGCGG 1534767
QY 526 GTTCCGGCTTCGCTGTCGCGCGCTCAGATGTGGGACAGCGTGGGAGTACCTGTTT 585
Db 1534766 GTTCCGGCTTCGCTGTCGCGCGCTCAGATGTGGGACAGCGTGGGAGTACCTGTTT 1534707
QY 586 CGGCGCGCTTCGCGCGTTTCACTCGGTGCTCTGGGCTCTGACGCGGATCGTGGATAGTT 645
Db 1534706 CGGCGCGCTTCGCGCGTTTCACTCGGTGCTCTGGGCTCTGACGCGGATCGTGGATAGTT 1534647
QY 646 CGTCCGCGCTTCGATGTCGCGCGCGCTCGCGTATGTGGCTGATGAGCGTCAACG 705
Db 1534646 CGTCCGCGCTTCGATGTCGCGCGCGCTCGCGTATGTGGCTGATGAGCGTCAACG 1534587
QY 706 CGGCGCGCGCTTCGATGTCGCGCGCGCTCGCGTATGTGGCTGATGAGCGTCAACG 765
Db 1534586 CGGCGCGCGCTTCGATGTCGCGCGCGCTCGCGTATGTGGCTGATGAGCGTCAACG 1534527
QY 766 CGTATGGGCTGACGCTGTCGCGCGCGCTCGCGTATGTGGCTGATGAGCGTCAACG 825
Db 1534526 CGTATGGGCTGACGCTGTCGCGCGCGCTCGCGTATGTGGCTGATGAGCGTCAACG 1534467
QY 826 TGATAGCGACCAACTCTTTGGGCAAAACACCCCGCGATCGCGTCAACGAGCGCGAAT 885
Db 1534466 TGATAGCGACCAACTCTTTGGGCAAAACACCCCGCGATCGCGTCAACGAGCGCGAAT 1534407
QY 886 ACGGCGAGATGTGGGCGCAACGCGCGCGCGCGATGTTTGGCTACGCGCGCGCGCGCGA 945
Db 1534406 ACGGCGAGATGTGGGCGCAACGCGCGCGCGCGATGTTTGGCTACGCGCGCGCGCGA 1534347
QY 946 CGGCGACGCGCGCTGTCGCTGCGGTCGAGGAGGCGCGGAGATGACGAGCGCGGTCGGC 1005
Db 1534346 CGGCGACGCGCGCTGTCGCTGCGGTCGAGGAGGCGCGGAGATGACGAGCGCGGTCGGC 1534287
QY 1006 TCCTCGAGCAGCGCGCGCGCTCGAGGAGGCGCTCGACACCGCGCGCGCGCGCGA 1065
Db 1534286 TCCTCGAGCAGCGCGCTCGAGGAGGCGCGGAGGCGCGCGCGCGCGCGCGA 1534227
QY 1066 TGACAAATGTGCGCGCGCGCTGCAACAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCT 1125
Db 1534226 TGACAAATGTGCGCGCGCGCTGCAACAGCTGCGCGCGCGCGCGCGCGCGCGCGCGCT 1534167
QY 1126 CTTCCAAAGCTGGGTGGCGCTGTGGAGAGCGGCTCTCGCGCGCGCGCGCGCGCGCGCG 1185
Db 1534166 TCGACCAACTGAGTGAATCTTGGAAAGCGCATCTCGCGCGCGCGCGCGCGCGCGCG 1534107
QY 1186 TGGTGTGATGGCGCAACCAACATGTCGATGACCACTCGGGGTGTGCGATGACCGCA 1245
Db 1534106 TGGTGTGATGTCGCAACCAACCGTGTGATGACCACTCGGGGTGTGCGATGACCGCA 1534047
QY 1246 CTTTCAGCTCGATGTTGAAGGCTTTGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1305
Db 1534046 CTTTCGACTCGATGTTGAAGGCTTTGCTCGCGCGCGCGCGCGCGCGCGCGCGCG 1533990
QY 1306 CGGCGCAAAACGGGCTCGCGCGCGATGAGCTCGCTGGGCGAGC-----TCGCTGG 1353
Db 1533989 CGGCGCAAAACGGGCTCGCGCGCGATGAGCTCGCTGGGCGAGCAGCTCGCTGGCTGG 1533930

QY 1354 GTTCTTCGGTCTGGCGGTGGGGTGGCGGCAACTTTGGGTGGCGGCTCGGTGCGTT 1413
Db 1533929 GTTCTTCGGTCTGGCGGTGGGGTGGCGGCAACTTTGGGTGGCGGCTCGGTGCGTT 1533870
QY 1414 CGTTCGCTGCGCGAGGCTGGCGCGGCGGCAACCCAGCAGTCAACCCCGGCGCGCGG 1473
Db 1533869 CGTTCGCTGCGCGAGGCTGGCGCGGCGGCAACCCAGCAGTCAACCCCGGCGCGCGG 1533810
QY 1474 CGTTCGCTGCGCGAGGCTGGCGCGGCGGCAACCCAGCAGTCAACCCCGGCGCGCGG 1533
Db 1533809 CGTTCGCTGCGCGAGGCTGGCGCGGCGGCAACCCAGCAGTCAACCCCGGCGCGCGG 1533750
QY 1534 GGTTCGCTGCGGCGAGTGGCGCGGCGGCGGCGG-----TGGTGGGCTCAGTGTGTGC 1587
Db 1533749 GGTTCGCTGCGGCGAGTGGCGCGGCGGCGGCGGCGGCGGTTAGCAATGCT 1533690
QY 1588 TGGTGTTCGCGCGAGGCTGATGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1536
Db 1533689 TGGTGTTCGCGCGAGGCTGATGTCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1533641

RESULT 13
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: PRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Query Match 41.2%; Score 942.6; DB 3; Length 4411529;
Best Local Similarity 89.1%; Pred. No. 1.5e-178;
Matches 1059; Conservative 0; Mismatches 109; Indels 21; Gaps 3;
QY 466 TGGTGGATTCGGGGGCTTACCAACCGGAGATCACTCCGAGGATGTACGCGGCGCGG 525
Db 1533630 TGGTGGATTCGGGGGCTTACCAACCGGAGATCACTCCGAGGATGTACGCGGCGCGG 1533571
QY 526 GTTCGGGCTCGCTGGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTGAACCTGTTT 585
Db 1533570 GTTCGGGCTCGCTGGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTGAACCTGTTT 1533511
QY 586 CGGCGCGGCTCGGCTTTCAGTCCGTGTCTGGGGTCTGACGTTGGGGTCTGATAGGTT 645
Db 1533510 CGGCGCGGCTCGGCTTTCAGTCCGTGTCTGGGGTCTGACGTTGGGGTCTGATAGGTT 1533451
QY 646 CGTTCGGGCTGATGTTGGCGGCGGCTTCGCGCTATGCGGCTGATGCGGCTGATGCGGCT 705
Db 1533450 CGTTCGGGCTGATGTTGGCGGCGGCTTCGCGCTATGCGGCTGATGCGGCTGATGCGGCT 1533391
QY 706 CGGCGGAGGCGAGTGAACCGCGCGCGGCTTCGCGGCTGATGCGGCTGATGCGGCTGATGCGGCT 765
Db 1533390 CGGCGGAGGCGAGTGAACCGCGCGCGGCTTCGCGGCTGATGCGGCTGATGCGGCTGATGCGGCT 1533331
QY 766 CGTATGGGCTGAGGTCGCGGCGGCTGATGCGGCGGAGACCGTGTGCTGCTGATGCTGCTGCTGCT 825
Db 1533330 CGTATGGGCTGAGGTCGCGGCGGCTGATGCGGCGGAGACCGTGTGCTGCTGATGCTGCTGCTGCT 1533271
QY 826 TGATAGGACCAACTCTTGGGGCAAAACACCCCGGCGATTCGCGGTCAACGAGGCGCGAAT 885

Db 1533270 TGATAGGACCAACTCTTGGGGCAAAACACCCCGGCGATTCGCGGTCAACGAGGCGCGAAT 1533211
QY 886 AGCGGAGATGTGGGCCCAAGACCGCCCGCGGATTTGGCTACGCGCGCGGCGGCGGCGG 945
Db 1533210 AGCGGAGATGTGGGCCCAAGACCGCCCGCGGATTTGGCTACGCGCGCGGCGGCGGCGG 1533151
QY 946 CGGCGACGGCGACGTTCTGCTCGCTTCAGGAGGCGCGGAGATGACACAGCGCGGCGGCGGCGG 1005
Db 1533150 CGGCGACGGCGAGGCTGCTGCTCGCTTCAGGAGACCGCCACTGATCAACACCCCGGCGGCGG 1533091
QY 1006 TCCTCGAGAGGCGCGCGCGGTCAGAGGCTTCGACACCGCGCGGCGGCGGCGGCGGCGGCGG 1065
Db 1533090 TCCTTGAGAGGCGCGCGCGGTCAGAGGCTTCGACACCGCGCGGCGGCGGCGGCGGCGGCGG 1533031
QY 1066 TGAACAATGTGCCCCAGCGCTGCAACAGCTGGGCCACCGCCACGACGAGGCGACCAACGCGCTT 1125
Db 1533030 TGAACAATGTGCCCCAGCGCTGCAACAACTGGGCCACCGCCACGAAAGCATCTGGCGGT 1532971
QY 1126 CTTCCAGCTGGGTGGCTTCGGAAGACGCTTCGCGCGATCGGTTCGCCGATCAGCAACA 1185
Db 1532970 TCGACCAACTGAGTGAACCTCTGGAAGGCCATCTCGCGCATCTGTCCGCGCTCAGCAACA 1532911
QY 1186 TGGTGTGATGGCCCAACCAACATGTCGATGACCAACTCGGGTGTCTCGATGACCAACA 1245
Db 1532910 TCGTGTGATGCTCAACCAACATGTCGATGACCAACTCGGGTGTCTCGATGACCAACA 1532851
QY 1246 CTTTGAGCTCGATGTTGAAGGCTTTTCTCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1305
Db 1532850 CTTTGCACTCGATGTTGAAGGCTTTTCTCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1532794
QY 1306 CGGCGCAAAACGGGCTCGGCGGATGATGCTCGTGGCGAGC-----TCGCTGG 1353
Db 1532793 CGGCGCAAAACGGGCTCGGCGGATGATGCTCGTGGCGAGCAGCTGGGTTCTGCTGCTGG 1532734
QY 1354 GTTCTTCGGGCTCGGCGGCTGGCGGCTGGCGCAACTTTGGGTTCGGCGGCGGCTTCGCTCGGTT 1413
Db 1532733 GTTCTTCGGGCTCGGCGGCTGGCGGCTGGCGCAACTTTGGGTTCGGCGGCGGCTTCGCTCGGTT 1532674
QY 1414 GGTTCGCTGCGCGAGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1473
Db 1532673 GGTTCGCTGCGCGAGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1532614
QY 1474 CGTTCGCTGACGAGCTGACGAGCGCGGCGGAGAGGCGGCGGCGGCGGCGGCGGCGGCGG 1533
Db 1532613 CGTTCGCTGACGAGCTGACGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1532554
QY 1534 GGTTCGCTGCGCGAGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1587
Db 1532553 GGTTCGCTGCGCGAGGCTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1532494
QY 1588 TGGTGTTCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1636
Db 1532493 TGGGATCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1532445

RESULT 14
US-08-818-112-110
; Sequence 110, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/818,112

FILING DATE: 13-MAR-1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.411C6

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 110:

SEQUENCE CHARACTERISTICS:

LENGTH: 3027 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-818-112-110

Query Match

Best Local Similarity 41.1%; Score 941; DB 3; Length 3027;

Mismatches 1058; Conservative 0; Mismatches 110; Indels 21; Gaps 3;

466 TGTGTGATTTCCGGGCGGTACACCGGAGATCAACTCCGGGAGATGACCGCGGCCCGG 525
 Db 1794 TGGTGGACTTCGGGGCGGTACACCGGAGATCAACTCCGGGAGATGATACCGCGGCCCGG 1853
 Qy 526 GTTCGGCTCGCTGGTGGCGCGCTCAGATGTGGGACAGCGTGGGAGTACTGTTTT 585
 Db 1854 GTTCGGCTCGCTGGTGGCGCGCTCAGATGTGGGACAGCGTGGGAGTACTGTTTT 1913
 Qy 586 CGGCGCGCTCGCTGGTGGCGCGCTCAGATGTGGGACAGCGTGGGAGTACTGTTTT 645
 Db 1914 CGGCGCGCTCGCTGGTGGCGCGCTCAGATGTGGGACAGCGTGGGAGTACTGTTTT 1973
 Qy 646 CGTGGCGGCTCGATGCTGGCGCGCTCGCGCTGATGTGGGAGTACTGTTTT 705
 Db 1974 CGTGGCGGCTCGATGCTGGCGCGCTCGCGCTGATGTGGGAGTACTGTTTT 2033
 Qy 706 CGGCGGAGCGGAGCTGACCGCGCGCTCGCGCTGATGTGGGAGTACTGTTTT 765
 Db 2034 CGGCGGAGCGGAGCTGACCGCGCGCTCGCGCTGATGTGGGAGTACTGTTTT 2093
 Qy 766 CGTATGGGCTGACGCTGCGCGCGCTGATGTGGGAGTACTGTTTT 825
 Db 2094 CGTATGGGCTGACGCTGCGCGCGCTGATGTGGGAGTACTGTTTT 2153
 Qy 826 TGATAGCGACCACTCTTGGGCGCAACACCGCGGATCGCGGTCAACGAGCGCGAT 885
 Db 2154 TGATAGCGACCACTCTTGGGCGCAACACCGCGGATCGCGGTCAACGAGCGCGAT 2213
 Qy 886 ACGGCGAGATGTGGGCGCAACGCGCGCGATGTTTGGCTACGCGCGCGCGCGCA 945
 Db 2214 ACGGGAGATGTGGGCGCAACGCGCGCGATGTTTGGCTACGCGCGCGCGCGCA 2273
 Qy 946 CGGCGAGCGGAGCTTCTCGCTTCCAGGAGCGCGCGAGATGACGAGCGCGGTGGC 1005
 Db 2274 CGGCGAGCGGAGCTTCTCGCTTCCAGGAGCGCGCGAGATGACGAGCGCGGTGGC 2333
 Qy 1006 TCCTCGAGCGCGCGCGGTGAGAGCGCTTCGACACCGCGCGCGGACCACTTGA 1065
 Db 2334 TCCTTGAGCGCGCGGTGAGAGCGCTTCGACACCGCGCGCGGACCACTTGA 2393

Qy 1066 TGAACAATGTGCCCCAGGCGCTGCAACAGCTGGGCCAGCCACGAGGCGACCGCCTT 1125
 Db 2394 TGAACAATGTGCCCCAGGCGCTGCAACAGCTGGGCCAGCCACGAGGCGATCTGGCGGT 2453
 Qy 1126 CTTCAAGCTGGTGGCTGTGGAGAGCGTCTCGCGCATCGGTTCGCGGATCAGCACCA 1185
 Db 2454 TCGACCAACTGAGTGAATCTGGAAGCCATCTCGCGCATCTGTGCGCGCTCAGCAACA 2513
 Qy 1186 TGGTGTGATGCGCAACCAACCAACATGTCGATGACCAACTCGGGTGTGTCGATCAGCAACA 1245
 Db 2514 TGGTGTGATGTCGCAACCAACCAACATGTCGATGACCAACTCGGGTGTGTCGATCAGCAACA 2573
 Qy 1246 CTTGAGCTCGATGTTGAAGGCTTTCCTCGGCGGCGCGCGCGAGCGCTGCAACCG 1305
 Db 2574 CTTGAGCTCAATGTTGAAGGCTTTCCTCGGCGGCGCGCGCGAGCGCTGCAACCG 2630
 Qy 1306 CGGCGCAAAACGGGCTCGCGCGATGAGCTCGCTGGGCGAGC-----TCGCTGG 1353
 Db 2631 CGGCGCAAAACGGGCTCGCGCGATGAGCTCGCTGGGCGAGC-----TCGCTGG 2690
 Qy 1354 GTTCTTGGGCTGCGGCGGTGGCGGCAACTTGGTGGGCGGCGCTCGGTGCGGTT 1413
 Db 2691 GTTCTTGGGCTGCGGCGGTGGCGGCAACTTGGTGGGCGGCGCTCGGTGCGGTT 2750
 Qy 1414 GCTTGTGGTGGCGGCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1473
 Db 2751 GCTTGTGGTGGCGGCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2810
 Qy 1474 GCTTGTGGTGGCGGCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1533
 Db 2811 GCTTGTGGTGGCGGCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2870
 Qy 1534 GCTTGTGGTGGCGGCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1587
 Db 2871 GCTTGTGGTGGCGGCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2930
 Qy 1588 TGGTGTGGTGGCGGCGGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1636
 Db 2931 TGGGATGCGCGCGGCGGCGGTACGTAATGCGCGCGGCGGCGGCGGCGGCGG 2979

RESULT 15

US-08-818-111-105

Sequence 105, Application US/08818111

Patent No. 6338852

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.

APPLICANT: Skelky, Yasir A.W.

APPLICANT: Dillon, Davin C.

APPLICANT: Campos-Neto, Antonia

APPLICANT: Houghton, Raymond

APPLICANT: Vedvick, Thomas S.

APPLICANT: Twardzik, Daniel R.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF

NUMBER OF SEQUENCES: 148

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED AND BERRY LLP

CITY: Seattle

STATE: Washington

COUNTRY: USA

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA: US/08/818,111

APPLICATION NUMBER: 424

FILING DATE: 13-MAR-1997

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.

TUBERCULOSIS

REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 105:
SEQUENCE CHARACTERISTICS:
LENGTH: 3027 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-105

Query Match 41.1%; Score 941; DB 4; Length 3027;
Best Local Similarity 89.0%; Pred. No. 9.4e-179;
Matches 1058; Conservative 0; Mismatches 110; Indels 21; Gaps 3;

466 TGGTGGATTTCGGGGCTTACACCGGAGATCAACTCCGCGAGGATGTACCGCGCCCGG 525
1794 TGGTGGACTTCGGGGCTTACACCGGAGATCAACTCCGCGAGGATGTACCGCGCCCGG 1853
526 GTTCGGCTCGCTGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTGAACCTGTTTT 585
1854 GTTCGGCTCGCTGGCGCGGCTGAGATGTGGGACAGCGTGGCGAGTGAACCTGTTTT 1913
586 CGCGCGCTCGCGCTTCACTCGGTGCTGGGGTCTGACCGTGGGGTCTGGATAGTTT 645
1914 CGCGCGCTCGCGCTTCACTCGGTGCTGGGGTCTGACCGAGGATCGTGGATAGTTT 1973
646 CGTCGGCGCTCGATGTGGCGCGGCTCGCGCTGATGTGGCGTGGATGACGTCACCG 705
1974 CGTCGGCGCTCGATGTGGCGCGGCTCGCGCTGATGTGGCGTGGATGACGTCACCG 2033
706 CGGGGACGCGGAGCTGACCGCGCGGCTCGCGCTGCTGGCGCGCTTACGAGACGG 765
2034 CGGGGACGCGGAGCTGACCGCGCGGCTCGCGCTGCTGGCGCGCTTACGAGACGG 2093
766 CGTATGGGCTGACGTCGCGCGCGGCTCGCGGAGACCGTCTGACTGATGATTC 825
2094 CGTATGGGCTGACGTCGCGCGCGGCTCGCGGAGACCGTCTGACTGATGATTC 2153
826 TGATAGCAGCAACCTCTTGGGGCAAAACACCGCGCGGCTCGCGTCAACGAGGCGGAAT 885
2154 TGATAGCAGCAACCTCTTGGGGCAAAACACCGCGCGGCTCGCGTCAACGAGGCGGAAT 2213
886 ACGGCGAGATGTGGGCGCAAGACCGCGCGGCTGCTGGCTACCGCGCGGCGGCGGA 945
2214 ACGGCGAGATGTGGGCGCAAGACCGCGCGGCTGCTGGCTACCGCGCGGCGGCGGA 2273
946 CGGCGACGCGAGCTTGTGCTGCTGCGAGGAGCGCGCGGATGACCGCGGCTGGCG 1005
2274 CGGCGACGCGAGCTTGTGCTGCTGCGAGGAGCGCGCGGATGACCGCGGCTGGCG 2333
1006 TCCTCGAGAGCGCGCGGCTGAGGAGGCTTCCGACACCGCGCGGCGAAACCAAGTTGA 1065
2334 TCCTCGAGAGCGCGCGGCTGAGGAGGCTTCCGACACCGCGCGGCGAAACCAAGTTGA 2393
1066 TGAACAATGTGCGGAGGCTGCAACAGCTGGCGCGGCGGCGGCGGATCGAGCGCCCTT 1125
2394 TGAACAATGTGCGGAGGCTGCAACAGCTGGCGCGGCGGCGGCGGATCGAGCGCCCTT 2453
1126 CTTCAGAGTGGGTGGCTGTGAGAGCGGTCTCGCGCGATCGGTGCGCGATCAGCAACA 1185
2454 TCGACCAACTGAGTGAACCTGGAAGAGCCATCTCGCGGATCTGTCGCGGCTCAGCAACA 2513
1186 TGGTGTGATGGCGCAACCAACATGTCGATGACCAACTCGGGTGTGTCGATGACCAACA 1245
2514 TCGTGTGATGTCACCAACCAACATGTCGATGACCAACTCGGGTGTGTCGATGGCGGCA 2573
1246 CCTTGAGCTCGATGTTGAAGGGCTTGTCTCGGCGCGGCGGCGGCGGCGGCGGCGG 1305
2574 CCTTGCACTCAATGTTGAAGGGCTTGTCTCGGCGCGGCGGCGGCGGCGGCGGCGG 2630

1306 CGCGCAAAACCGGGTCCCGGGCGATGAGCTCGCTGGGCGAGC-----TCGCTGG 1353
2631 CGCGCAAAACCGGGTCCAGGGCGATGAGCTCGCTGGGCGAGCAGCTGGGTTTCGTCGCTGG 2690
1354 GTTCCTCGGGTCTGGGCGGCTGGGGTGGCGCAACTTGGTTCGGGCGGCTCGCTCGGTT 1413
2691 GTTCCTCGGGTCTGGGCGGCTGGGGTGGCGCAACTTGGTTCGGGCGGCTCGCTCGGTT 2750
1414 CGTTGTCTGGTTCGCGCAGGCGCTGGGCGCGCCCAACCCAGGAGTACCCCGGCGGCGCGG 1473
2751 CGTTGTCTGGTTCGCGCAGGCGCTGGGCGCGCCCAACCCAGGCGTACCCCGGCGGCGCGG 2810
1474 CGCTGCGCTGACAGCCTGACCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1533
2811 CGCTGCGCTGACAGCCTGACCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2870
1534 GGCTGCGGCTGGGCGAGATGGGCGCGCAGGCGCGG-----TGGTGGGCTCAGTGGTGTGC 1587
2871 GGCTACCGCTGGGCGCAACTGACCAATAGCGGCGGCGGTTTCGGCGGGTTAGCAATGCGT 2930
1588 TGGTGTTCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1636
2931 TGGGATGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2979

Search completed: July 3, 2004, 22:00:26
Job time : 207 secs

1. The first part of the document is a letter from the President of the United States to the Congress, dated January 3, 1862. It is a very long letter, and it contains a great deal of information about the state of the country at that time. The President talks about the war with Mexico, and about the situation in the South. He also talks about the economy, and about the need for more money. The letter is written in a very formal style, and it is very long. It is a very important document, and it is one of the most important documents in the history of the United States.

2. The second part of the document is a letter from the Secretary of the Treasury to the Congress, dated January 3, 1862. It is a very long letter, and it contains a great deal of information about the state of the country at that time. The Secretary talks about the war with Mexico, and about the situation in the South. He also talks about the economy, and about the need for more money. The letter is written in a very formal style, and it is very long. It is a very important document, and it is one of the most important documents in the history of the United States.

3. The third part of the document is a letter from the Secretary of the Treasury to the Congress, dated January 3, 1862. It is a very long letter, and it contains a great deal of information about the state of the country at that time. The Secretary talks about the war with Mexico, and about the situation in the South. He also talks about the economy, and about the need for more money. The letter is written in a very formal style, and it is very long. It is a very important document, and it is one of the most important documents in the history of the United States.

4. The fourth part of the document is a letter from the Secretary of the Treasury to the Congress, dated January 3, 1862. It is a very long letter, and it contains a great deal of information about the state of the country at that time. The Secretary talks about the war with Mexico, and about the situation in the South. He also talks about the economy, and about the need for more money. The letter is written in a very formal style, and it is very long. It is a very important document, and it is one of the most important documents in the history of the United States.

5. The fifth part of the document is a letter from the Secretary of the Treasury to the Congress, dated January 3, 1862. It is a very long letter, and it contains a great deal of information about the state of the country at that time. The Secretary talks about the war with Mexico, and about the situation in the South. He also talks about the economy, and about the need for more money. The letter is written in a very formal style, and it is very long. It is a very important document, and it is one of the most important documents in the history of the United States.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2004, 20:19:36 ; Search time 1007 Seconds
(without alignments)
10956.269 Million cell updates/sec

Title: US-09-597-796c-11

Perfect score: 2287

Sequence: 1 tctagaataatttggttta.....ggntgttaaacagcccgaaa 2287

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 3163042 seqs, 2412103800 residues

Total number of hits satisfying chosen parameters: 6326084

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/prodata/2/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/prodata/2/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/prodata/2/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/prodata/2/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/prodata/2/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/prodata/2/pubpna/PCT05_PUBCOMB.seq.*
- 7: /cgn2_6/prodata/2/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/prodata/2/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/prodata/2/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/prodata/2/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/prodata/2/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/prodata/2/pubpna/US09A_PUBCOMB.seq.*
- 14: /cgn2_6/prodata/2/pubpna/US10A_PUBCOMB.seq.*
- 15: /cgn2_6/prodata/2/pubpna/US10B_PUBCOMB.seq.*
- 16: /cgn2_6/prodata/2/pubpna/US10C_PUBCOMB.seq.*
- 17: /cgn2_6/prodata/2/pubpna/US10_NEW_PUB.seq.*
- 18: /cgn2_6/prodata/2/pubpna/US60_NEW_PUB.seq.*
- 19: /cgn2_6/prodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2284	99.9	2287	9	US-09-287-849-1
2	2284	99.9	2287	12	US-09-886-349A-15
3	2284	99.9	2287	15	US-10-359-460-1
4	2284	99.9	2287	15	US-10-098-732A-15
5	2284	99.9	2287	16	US-10-359-459-1
6	2186.8	95.6	2451	16	US-10-369-983-8
7	2186.8	95.6	2487	16	US-10-369-983-7
8	2186.8	95.6	2637	16	US-10-369-983-6
9	2186.8	95.6	2808	15	US-10-098-732A-64
10	2186.8	95.6	2808	16	US-10-369-983-5
11	2186.8	95.6	3060	16	US-10-369-983-11
12	2186.8	95.6	3104	16	US-10-369-983-10
13	2186.8	95.6	3474	16	US-10-369-983-9
14	2185.2	95.5	2190	12	US-09-886-349A-17

15	2185.2	95.5	2190	15	US-10-098-732A-17	Sequence 17, Appl
16	1770.8	77.4	1797	9	US-09-287-849-25	Sequence 25, Appl
17	1770.8	77.4	1797	12	US-09-886-349A-19	Sequence 19, Appl
18	1770.8	77.4	1797	15	US-10-359-460-25	Sequence 25, Appl
19	1770.8	77.4	1797	15	US-10-098-732A-19	Sequence 19, Appl
20	1582	69.2	3030	16	US-10-369-983-3	Sequence 3, Appl
21	1572.6	68.8	2181	16	US-10-369-983-1	Sequence 1, Appl
22	1172.8	51.3	1801	9	US-09-287-849-21	Sequence 21, Appl
23	1172.8	51.3	1801	15	US-10-359-460-21	Sequence 21, Appl
24	1169	51.1	3058	12	US-09-886-349A-13	Sequence 13, Appl
25	1169	51.1	3058	15	US-10-193-002-101	Sequence 106, Appl
26	1169	51.1	3058	15	US-10-084-843-106	Sequence 13, Appl
27	1169	51.1	75216	15	US-10-098-732A-13	Sequence 646, App
28	1169	51.1	75216	15	US-10-080-170-646	Sequence 646, App
29	1169	51.1	75216	17	US-10-080-170-646	Sequence 105, App
30	941	41.1	3027	15	US-10-193-002-105	Sequence 110, App
31	941	41.1	3027	15	US-10-084-843-110	Sequence 26271, A
32	921.8	40.3	1179	13	US-10-282-122A-25271	Sequence 28708, A
33	930.2	40.2	1182	13	US-10-282-122A-28708	Sequence 103, App
34	884	38.7	1725	15	US-10-193-002-103	Sequence 108, App
35	884	38.7	1725	15	US-10-084-843-108	Sequence 11, Appl
36	718.8	31.4	851	12	US-09-886-349A-11	Sequence 33, Appl
37	718.8	31.4	851	15	US-10-193-002-33	Sequence 33, Appl
38	718.8	31.4	851	15	US-10-084-843-33	Sequence 11, Appl
39	718.8	31.4	851	15	US-10-098-732A-11	Sequence 15, Appl
40	591.2	25.9	1068	9	US-09-712-363-15	Sequence 5, Appl
41	589	25.8	1002	12	US-09-886-349A-5	Sequence 5, Appl
42	589	25.8	1002	15	US-10-098-732A-5	Sequence 5, Appl
43	588	25.7	1872	12	US-09-886-349A-1	Sequence 17, Appl
44	588	25.7	1872	15	US-10-193-002-17	Sequence 17, Appl
45	588	25.7	1872	15	US-10-084-843-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1

US-09-287-849-1
; Sequence 1, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillion, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2287
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein Ra12-TbH9-Ra35 (designated Mtb32-MCb39)
; OTHER INFORMATION: fusion)
; NAME/KEY: modified_base

```

; LOCATION: (30)
; OTHER INFORMATION: n = g, a, c or t
; NAME/KEY: modified_base
; LOCATION: (33)
; OTHER INFORMATION: n = g, a, c or t
; NAME/KEY: CDS
; LOCATION: (42)..(2231)
; NAME/KEY: modified_base
; LOCATION: (2270)
; OTHER INFORMATION: n = g, a, c or t
US-09-287-849-1

Query Match      99.9%; Score 2284; DB 9; Length 2287;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2287; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTAGAAATAATTTTGTCTTAAAGAAAGANATATACATATCATCATCACCATCACCATC 60
DB 1 TCTAGAAATAATTTTGTCTTAAAGAAAGANATATACATATCATCATCACCATCACCATC 60

QY 61 ACACGGCCGCTCCGATTAATTCAGTGTCCAGGGTGGCCAGGATTCGCCATTCGGA 120
DB 61 ACACGGCCGCTCCGATTAATTCAGTGTCCAGGGTGGCCAGGATTCGCCATTCGGA 120

QY 121 TCGGSCAGCGATGCGGATCGGGCCAGATCCGATCGGGTGGGGGTCAACCAACCGTTC 180
DB 121 TCGGSCAGCGATGCGGATCGGGCCAGATCCGATCGGGTGGGGGTCAACCAACCGTTC 180

QY 181 ATATCGGCGCTACCGCTTCTCGCTTGGTGTGTCGACAAACAGCGCAACCGGCGCAC 240
DB 181 ATATCGGCGCTACCGCTTCTCGCTTGGTGTGTCGACAAACAGCGCAACCGGCGCAC 240

QY 241 GAGTCCAAACGCTGGTGGGAGCGCTCGGGCGGAAGTCTCGGATCTCCACGGCGAAG 300
DB 241 GAGTCCAAACGCTGGTGGGAGCGCTCGGGCGGAAGTCTCGGATCTCCACGGCGAAG 300

QY 301 TGATCACCGCGGTGCAAGCGGCTCCGATCAACTCGGCGCACCGGATGGCGGAGCGGCTTA 360
DB 301 TGATCACCGCGGTGCAAGCGGCTCCGATCAACTCGGCGCACCGGATGGCGGAGCGGCTTA 360

QY 361 ACGGGCAATATCCGGTGAAGTCACTCGGTGACCTGGGAAAACAGTCCGGCGGCAACG 420
DB 361 ACGGGCAATATCCGGTGAAGTCACTCGGTGACCTGGGAAAACAGTCCGGCGGCAACG 420

QY 421 GTACAGGGAACGTGACATTCGCGAGGACCCCGCGCGAATTCATGTTGGATTTGGGG 480
DB 421 GTACAGGGAACGTGACATTCGCGAGGACCCCGCGCGAATTCATGTTGGATTTGGGG 480

QY 481 GTTACACCGGAGATCAATCCGCGAGGATGTAACCGCGCGCGGTTTCGGGCTCGCTGG 540
DB 481 GTTACACCGGAGATCAATCCGCGAGGATGTAACCGCGCGCGGTTTCGGGCTCGCTGG 540

QY 541 TGGCCGGGCTCAGATGCGGAGCAGGTGCGGAGTGACCTGTTTCGGCGCGGCTCGGCT 600
DB 541 TGGCCGGGCTCAGATGCGGAGCAGGTGCGGAGTGACCTGTTTCGGCGCGGCTCGGCT 600

QY 601 TTCAGTCGGTGTGGGGTCTGACGCTGGGGTCTGGGATAGGTTCTGTCGGGGTCTGA 660
DB 601 TTCAGTCGGTGTGGGGTCTGACGCTGGGGTCTGGGATAGGTTCTGTCGGGGTCTGA 660

QY 661 TTGTTGGGGGGGCTCGCCGTATGTGGCTGGATGAGCGTACCGGGGGGCGAGCGGAGC 720
DB 661 TTGTTGGGGGGGCTCGCCGTATGTGGCTGGATGAGCGTACCGGGGGGCGAGCGGAGC 720

QY 721 TGACCGCGCGCCAGGTTCGGGTTGCTGCGCGCGCTACGAGACGGCGTATGGGCTGACGG 780
DB 721 TGACCGCGCGCCAGGTTCGGGTTGCTGCGCGCGCTACGAGACGGCGTATGGGCTGACGG 780

QY 781 TGCCCCCGCGGTGATCGCGGAGAACCGTGTGAACTGATGATTTGATGAGGACCAACC 840
DB 781 TGCCCCCGCGGTGATCGCGGAGAACCGTGTGAACTGATGATTTGATGAGGACCAACC 840

QY 841 TCTTGGGGGAAAAACACCCCGGGGATCGCGGTCAAACGAGCGCCGAATACCGCGAGATGTGG 900

```


Db 1921 AGGATGTCGGGTGCTGAGCTGCGCGGTGCGGTGCGCTGCGTGGCGGCGATCGGTG 1980
QY 1981 GCGCGTCCGGTGGTGGAGCCCGTGGTGGCGATCGGCGAAAGCGGTGGCGGCGGAA 2040
Db 1981 GCGCGTCCGGTGGTGGAGCCCGTGGTGGCGATCGGCGAAAGCGGTGGCGGCGGAA 2040
QY 2041 CGCCCGTCCGGTGGTGGAGCCCGTGGTGGCGATCGGCGAAAGCGGTGGCGGCGGAA 2100
Db 2041 CGCCCGTCCGGTGGTGGAGCCCGTGGTGGCGATCGGCGAAAGCGGTGGCGGCGGAA 2100
QY 2101 CGGTGACCGGTGCGGAGAGACATTGAACGGGTGGTGGATCGGCGGCGGATCGGCGG 2160
Db 2101 CGGTGACCGGTGCGGAGAGACATTGAACGGGTGGTGGATCGGCGGCGGATCGGCGG 2160
QY 2161 CGGTGATTCGGCGCGGCGGCGGCGTGGTGGAGCGGTGGTGGAGCGGTGGTGGAGCGG 2220
Db 2161 CGGTGATTCGGCGCGGCGGCGGCGTGGTGGAGCGGTGGTGGAGCGGTGGTGGAGCGG 2220
QY 2221 CGCGTCTAGGATATCCATCATCTGCGGCGGCGGTGGAGCGATCGCGGCGGATCGG 2280
Db 2221 CGCGTCTAGGATATCCATCATCTGCGGCGGCGGTGGAGCGATCGCGGCGGATCGG 2280
QY 2281 CCGGAAA 2287
Db 2281 CCGGAAA 2287

RESULT 2
US-09-886-349A-15
; Sequence 15, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Steven
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-0097005
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 2287
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein MTB72F (Ra12-TbH9-Ra35 or MTB32-MTB39
; OTHER INFORMATION: fusion)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (42)..(2231)
; OTHER INFORMATION: MTB72F
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (30)
; OTHER INFORMATION: n = g, a, c or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (33)
; OTHER INFORMATION: n = g, a, c or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2270)
; OTHER INFORMATION: n = g, a, c or t
; US-09-886-349A-15

Query Match

99.9%; Score 2284; DB 12; Length 2287;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 2287; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCTAGAATAATTTGTTTACTTTTAAAGANGANATATACATATGCAATCACCATCACCATC 60
Db 1 TCTAGAATAATTTGTTTACTTTTAAAGANGANATATACATATGCAATCACCATCACCATC 60
QY 61 ACACGGCGCGGTCCGATAAATTCACGCTGTCCAGGGTGGCGAGGGATTCGCAATTCGGA 120
Db 61 ACACGGCGCGGTCCGATAAATTCACGCTGTCCAGGGTGGCGAGGGATTCGCAATTCGGA 120
QY 121 TCGGGAGGCGATGCGCGATCGCGGCGCAGATCGATCGGTCGGGTGGGGTCCACCAACGCTTC 180
Db 121 TCGGGAGGCGATGCGCGATCGCGGCGCAGATCGATCGGTCGGGTGGGGTCCACCAACGCTTC 180
QY 181 ATATCGGGCGCTTACCGCTCTCTCGGCTTGGGTGTTCGACAAACACGCGCAACGCGCGAC 240
Db 181 ATATCGGGCGCTTACCGCTCTCTCGGCTTGGGTGTTCGACAAACACGCGCAACGCGCGAC 240
QY 241 GAGTCCAAACGCGTGGTGGAGCGCTCCGGGCGCAAGTCTCGGATCTCCACCGGCGACG 300
Db 241 GAGTCCAAACGCGTGGTGGAGCGCTCCGGGCGCAAGTCTCGGATCTCCACCGGCGACG 300
QY 301 TGATCACCGCGGTGACGCGCTCCGATCAACTCGGCGCACCGGATGGCGGACGCGCTTA 360
Db 301 TGATCACCGCGGTGACGCGCTCCGATCAACTCGGCGCACCGGATGGCGGACGCGCTTA 360
QY 361 ACGGGCGATCATCCCGGTGACGTCTCTCGGTGACCTCGGTGACCTCGGTGACCTCGGTG 420
Db 361 ACGGGCGATCATCCCGGTGACGTCTCTCGGTGACCTCGGTGACCTCGGTGACCTCGGTG 420
QY 421 GTACAGGAGACGTGACATTTGGCGGAGGAGCCCGCGGCGAATTCATGTTGATTTCCGGG 480
Db 421 GTACAGGAGACGTGACATTTGGCGGAGGAGCCCGCGGCGAATTCATGTTGATTTCCGGG 480
QY 481 CGTTACACACGCGGATCAACTCCGCGAGGATGTACGCGCGCGCGGCTTCGCGCTCGCTGG 540
Db 481 CGTTACACACGCGGATCAACTCCGCGAGGATGTACGCGCGCGCGGCTTCGCGCTCGCTGG 540
QY 541 TGGCGCGCGCTCAGATGTTGGGAGAGCGTGGCGAGTACCTGTTTTCGCGCGCGTCCGGT 600
Db 541 TGGCGCGCGCTCAGATGTTGGGAGAGCGTGGCGAGTACCTGTTTTCGCGCGCGTCCGGT 600
QY 601 TTCAGTTCGCTGCTTCGGGTCTGACCGGTGGGTCTGTGATAGGTTCGTTCGCGCGGTCTGA 660
Db 601 TTCAGTTCGCTGCTTCGGGTCTGACCGGTGGGTCTGTGATAGGTTCGTTCGCGCGGTCTGA 660
QY 661 TGGTGGCGCGCGCTTCGCGCTATGTCGCTGATGAGCGTCAACCGCGGCGAGCGCGAGC 720
Db 661 TGGTGGCGCGCGCTTCGCGCTATGTCGCTGATGAGCGTCAACCGCGGCGAGCGCGAGC 720
QY 721 TGACCGCGCGCGCTTCGCGGTTCGCGCGGCTTACGAGCGCGCTATGGGTGACGG 780
Db 721 TGACCGCGCGCGCTTCGCGGTTCGCGCGGCTTACGAGCGCGCTATGGGTGACGG 780
QY 781 TGCCCCCGCGCGGTGATCGCGGAGAACCGTGTGAACTGATGATTCGTAGCGACCAACC 840
Db 781 TGCCCCCGCGCGGTGATCGCGGAGAACCGTGTGAACTGATGATTCGTAGCGACCAACC 840
QY 841 TCTTGGGGCAAAACACCCCGCGGATCGCGGTCAACGAGCGCGAATACCGCGAGATGTGG 900
Db 841 TCTTGGGGCAAAACACCCCGCGGATCGCGGTCAACGAGCGCGAATACCGCGAGATGTGG 900
QY 901 CCCAAGACGCGCGCGGTGATGTTGGCTACCGCGCGCGACCGCGCGAGCGCGAGCGT 960
Db 901 CCCAAGACGCGCGCGGTGATGTTGGCTACCGCGCGCGACCGCGCGAGCGCGAGCGT 960
QY 961 TGCTGCGGTTCGAGGAGCGCGGAGATGACACGCGCGGTGGGTCTCTCGAGCGGCGG 1020
Db 961 TGCTGCGGTTCGAGGAGCGCGGAGATGACACGCGCGGTGGGTCTCTCGAGCGGCGG 1020
QY 1021 CCGCGGTTCGAGGAGCGCTCCGACACCGCGCGCGGCAACAGTTGATGAACATGTGCCCC 1080
Db 1021 CCGCGGTTCGAGGAGCGCTCCGACACCGCGCGCGGCAACAGTTGATGAACATGTGCCCC 1080

Db 1021 CGCGGTGAGAGAGCCTCCGACACCGCGCGCGGCGAACCAGTTGATGAACAATGTGCCCC 1080
QY 1081 AGGCGCTCAACAGCTGGCCAGCCACACGAGGACCAACAGCCCTTCTTCAAGCTGGGTG 1140
Db 1081 AGGCGCTCAACAGCTGGCCAGCCACACGAGGACCAACAGCCCTTCTTCAAGCTGGGTG 1140
QY 1141 GCCTGTGAAGACGCTCTCGCGCATCGGTGCCGATCAGCAACATGCTGTCATGACCA 1200
Db 1141 GCCTGTGAAGACGCTCTCGCGCATCGGTGCCGATCAGCAACATGCTGTCATGACCA 1200
QY 1201 ACAACCATGTCGATGACCAACTCGGGTGTGTGATGACCAACACCTTGAGCTCGATGT 1260
Db 1201 ACAACCATGTCGATGACCAACTCGGGTGTGTGATGACCAACACCTTGAGCTCGATGT 1260
QY 1261 TGAAGGGCTTTGCTCCGCGCGCGGCGCCAGCGCTGTCAGGCGGTGCAAAACCGGGG 1320
Db 1261 TGAAGGGCTTTGCTCCGCGCGCGGCGCCAGCGCTGTCAGGCGGTGCAAAACCGGGG 1320
QY 1321 TCCGGGCGATGAGCTCGCTGGGACAGCTCGCTGGGTCTTCTCGGCTCTCGGCGGTGGTGG 1380
Db 1321 TCCGGGCGATGAGCTCGCTGGGACAGCTCGCTGGGTCTTCTCGGCTCTCGGCGGTGGTGG 1380
QY 1381 CGGCCAACTTGGGTGGGCGGCTCGGTGCGTTCGTTGTGCTGCTGCCAGCGCTGCGCGG 1440
Db 1381 CGGCCAACTTGGGTGGGCGGCTCGGTGCGTTCGTTGTGCTGCTGCCAGCGCTGCGCGG 1440
QY 1441 CGGCCAACTGAGCTACACCGCGCGGCGGCGCTGCGCTGACAGCTGACAGCG 1500
Db 1441 CGGCCAACTGAGCTACACCGCGCGGCGGCGCTGCGCTGACAGCTGACAGCG 1500
QY 1501 CGCGGAAAGAGGCGCGGCGAGATGCTGGGCGGCTGCGGCTGGGCGAGATGGGCGCCA 1560
Db 1501 CGCGGAAAGAGGCGCGGCGAGATGCTGGGCGGCTGCGGCTGGGCGAGATGGGCGCCA 1560
QY 1561 GGGCGGTGGTGGGCTGAGTGGTGTGCTGGTTCGCGCGGACCTATGATGCGCG 1620
Db 1561 GGGCGGTGGTGGGCTGAGTGGTGTGCTGGTTCGCGCGGACCTATGATGCGCG 1620
QY 1621 ATTCTCGGAGCGCGGATATCCCGCGGCTTGTGCGAGGACCGGTTGCGCGACT 1680
Db 1621 ATTCTCGGAGCGCGGATATCCCGCGGCTTGTGCGAGGACCGGTTGCGCGACT 1680
QY 1681 TCCCGGCTGCGCTCGACCGCTCGCGATGCTGCGCAAGTGGGCGCACAGTGGTCA 1740
Db 1681 TCCCGGCTGCGCTCGACCGCTCGCGATGCTGCGCAAGTGGGCGCACAGTGGTCA 1740
QY 1741 ACATCAACCAACTGGCTTACAAACCGCTGGGCGCGGACCGGATCGTCATCG 1800
Db 1741 ACATCAACCAACTGGCTTACAAACCGCTGGGCGCGGACCGGATCGTCATCG 1800
QY 1801 ATCCCAACGCTGCTGCTGCTACCAACCAACAGCTGATCGCGGCGCACCATCAATG 1860
Db 1801 ATCCCAACGCTGCTGCTGCTACCAACCAACAGCTGATCGCGGCGCACCATCAATG 1860
QY 1861 CGTTTCAGCTGCGCTCGCGCAACCTTACCGCTCGATGCTGGTGGGATGACCGCACCC 1920
Db 1861 CGTTTCAGCTGCGCTCGCGCAACCTTACCGCTCGATGCTGGTGGGATGACCGCACCC 1920
QY 1921 AGGATGTCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
Db 1921 AGGATGTCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
QY 1981 GCGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040
Db 1981 GCGGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040
QY 2041 CGCCCGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2100
Db 2041 CGCCCGTGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2100
QY 2101 CGTGTACCGGTGCGGAGAGACATTGAACGGGTGATGATGATGATGATGATGATGATGATG 2160
Db 2101 CGTGTACCGGTGCGGAGAGACATTGAACGGGTGATGATGATGATGATGATGATGATGATG 2160

QY 2161 CGCGTGAATTCGGGCGGCGCGCTCGTCAACCGCTAGGACAGGTGGTTCGGTATGAACACGG 2220
Db 2161 CGCGTGAATTCGGGCGGCGCGCTCGTCAACCGCTAGGACAGGTGGTTCGGTATGAACACGG 2220
QY 2221 CGCGTCTTAGATATCCATCACTGCGGCGCGCTCGAGCAGATCCGGTGTAAACAAAG 2280
Db 2221 CGCGTCTTAGATATCCATCACTGCGGCGCGCTCGAGCAGATCCGGTGTAAACAAAG 2280
QY 2281 CCGGAAA 2287
Db 2281 CCGGAAA 2287

RESULT 3

US-10-359-460-1
; Sequence 1, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Devin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2287
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein Ra12-TbH9-Ra35 (designated MtB32-Mtb39
; OTHER INFORMATION: fusion)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (30)
; OTHER INFORMATION: n = g, a, c or t
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (33)
; OTHER INFORMATION: n = g, a, c or t
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (42)..(2231)
; FEATURE:
; NAME/KEY: modified_base
; LOCATION: (2270)
; OTHER INFORMATION: n = g, a, c or t
US-10-359-460-1

Query Match 99.9%; Score 2284; DB 15; Length 2287;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2287; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 2161 CCGGTGATTTCGGCGGGCCCGCTCGTCAACGCGCTAGACAGGTGGTATGAACACGG 2220
Qy 2221 CCGCGTCTTAGATATCATCAGTGGCGCGCTCGAGCAGATCCGNTGTAAACAG 2280
Db 2221 CCGCGTCTTAGATATCATCAGTGGCGCGCTCGAGCAGATCCGNTGTAAACAG 2280
Qy 2281 CCCGAAA 2287
Db 2281 CCCGAAA 2287

RESULT 4
US-10-098-732A-15
; Sequence 15, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 2287
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein MTB72F (Ra12-TbH9-Ra35 or MTB32-MTB39
; OTHER INFORMATION: fusion)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (42)..(2231)
; OTHER INFORMATION: MTB72F
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (1)..(2287)
; OTHER INFORMATION: n = g, a, c or t
US-10-098-732A-15

Query Match 99.9%; Score 2284; DB 15; Length 2287;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2287; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCTAGAATAATTTGTTTACTTTAAGAANGANATATACATATGCATCACCATCACCATC 60
Db 1 TCTAGAATAATTTGTTTACTTTAAGAANGANATATACATATGCATCACCATCACCATC 60

Qy 61 ACACGGCCGCTCCGATAAATTCAGCTGTCCAGAGGTGGCGAGGATTCGCCATTCGGA 120
Db 61 ACACGGCCGCTCCGATAAATTCAGCTGTCCAGAGGTGGCGAGGATTCGCCATTCGGA 120

Qy 121 TCGGGCAGGCGATGGCGATCGCGGCAGATCCGATCGGGTGGGGGTACCCACCGTTTC 180
Db 121 TCGGGCAGGCGATGGCGATCGCGGCAGATCCGATCGGGTGGGGGTACCCACCGTTTC 180

Qy 181 ATATCGGGCTACCGCTTCTCGGCTTGGGTGTTGTGCAACAACACGCGACCGGCGAC 240
Db 181 ATATCGGGCTACCGCTTCTCGGCTTGGGTGTTGTGCAACAACACGCGACCGGCGAC 240

Qy 241 GAGTCCAAACCGGTGGAGCGCTCCGCGCGCAAGTCTCGGCAATCTCCACCGGAGC 300
Db 241 GAGTCCAAACCGGTGGAGCGCTCCGCGCGCAAGTCTCGGCAATCTCCACCGGAGC 300

Qy 301 TGATCAACCGGGTCGACGGGCTCCGATCAACCTCGGCCACCGGATGGCGGACGCGTTA 360
Db 301 TGATCAACCGGGTCGACGGGCTCCGATCAACCTCGGCCACCGGATGGCGGACGCGTTA 360

Db 301 TGATCAACCGGGTCGACGGGCTCCGATCAACCTCGGCCACCGGATGGCGGACGCGTTA 360
Qy 361 ACGGGCATCATCCCGTGCAGTCTCTCGGTGACCTTCTGGTGAACCTGGCAACCAAGTCGGCGGACGC 420
Db 361 ACGGGCATCATCCCGTGCAGTCTCTCGGTGACCTTCTGGTGAACCTGGCAACCAAGTCGGCGGACGC 420
Qy 421 GTACAGGGAACGTGACATTTGGCCGAGGAGACCCCGCGCGCAATTCATGCTGGATTTTCGGGG 480
Db 421 GTACAGGGAACGTGACATTTGGCCGAGGAGACCCCGCGCGCAATTCATGCTGGATTTTCGGGG 480
Qy 481 CGTTACACCGGAGATCAACTCCGCGAGGATGTACGCGCGCCCGGGTTTCGCGCTCGCTGG 540
Db 481 CGTTACACCGGAGATCAACTCCGCGAGGATGTACGCGCGCCCGGGTTTCGCGCTCGCTGG 540
Qy 541 TGGCCGCGGCTCAGATGTGGGACACGCTGGCGAGTGACCTGTTTCGSCCGCTCGCGGT 600
Db 541 TGGCCGCGGCTCAGATGTGGGACACGCTGGCGAGTGACCTGTTTCGSCCGCTCGCGGT 600
Qy 601 TTCAGTCCGTGCTTGGGCTCTGACGCTGGGCTCGTGCATAGGTTTCGCGGGTCTGA 660
Db 601 TTCAGTCCGTGCTTGGGCTCTGACGCTGGGCTCGTGCATAGGTTTCGCGGGTCTGA 660
Qy 661 TGGTGGCGGCGGCTCGCGCTATGTGGCGTGAGGCTCACCGCGGGGAGCGGAGC 720
Db 661 TGGTGGCGGCGGCTCGCGCTATGTGGCGTGAGGCTCACCGCGGGGAGCGGAGC 720
Qy 721 TGACCGCGCGGCTCGCGGCTTGTGCGCGCGGCTACGAGACGCGCTATGGGCTGACGG 780
Db 721 TGACCGCGCGGCTCGCGGCTTGTGCGCGCGGCTACGAGACGCGCTATGGGCTGACGG 780
Qy 781 TGCCCCCGCGGCTGATTCGCCGAGAACCGTGTCTGAATGATGATTCGATAGCAGCAACC 840
Db 781 TGCCCCCGCGGCTGATTCGCCGAGAACCGTGTCTGAATGATGATTCGATAGCAGCAACC 840
Qy 841 TCTTGGGCAAAAACACCCCGCGGATTCGGGTCAACAGGCGCGAATACGCGGAGATGTGG 900
Db 841 TCTTGGGCAAAAACACCCCGCGGATTCGGGTCAACAGGCGCGAATACGCGGAGATGTGG 900
Qy 901 CCCAAGACGCGCGCGGATTTTGGCTACGCGCGGCGAAGCGCGACGCGCGACGT 960
Db 901 CCCAAGACGCGCGCGGATTTTGGCTACGCGCGGCGAAGCGCGACGCGCGACGT 960
Qy 961 TGCTGCGGTTTCGAGAGGCGCGGAGATGACACCGCGGGTGGGCTCTCGAGCAGCGC 1020
Db 961 TGCTGCGGTTTCGAGAGGCGCGGAGATGACACCGCGGGTGGGCTCTCGAGCAGCGC 1020
Qy 1021 CCGCGGTCGAGGAGGCGCTCCGACACCGCGCGGCGAACCAGTTGATGAACAATGTGCCCC 1080
Db 1021 CCGCGGTCGAGGAGGCGCTCCGACACCGCGCGGCGAACCAGTTGATGAACAATGTGCCCC 1080
Qy 1081 AGGCGCTGACACGCTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
Db 1081 AGGCGCTGACACGCTGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1140
Qy 1141 GCCTGTGGAAGACGCTCTCGCGCATCGGTCCGCGATCAGCAACATCGTTCGATGCGCCA 1200
Db 1141 GCCTGTGGAAGACGCTCTCGCGCATCGGTCCGCGATCAGCAACATCGTTCGATGCGCCA 1200
Qy 1201 ACAACCAATGTTCATGACCACTCGGCTGTGCTGATGACCAACCTTGAAGTCTGATGT 1260
Db 1201 ACAACCAATGTTCATGACCACTCGGCTGTGCTGATGACCAACCTTGAAGTCTGATGT 1260
Qy 1261 TGAAGGCTTTTGTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
Db 1261 TGAAGGCTTTTGTCCCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
Qy 1321 TCGGGCGATGAGCTCGCTGGGCGAGCTCGCTGGGTTCTTCGGGTCTGGCGGTGGGGTGG 1380
Db 1321 TCGGGCGATGAGCTCGCTGGGCGAGCTCGCTGGGTTCTTCGGGTCTGGCGGTGGGGTGG 1380
Qy 1381 CCGCCAACTTTGGGTTCGGGCGGCGCTCGGTTCGTTGTCGGTTCGGCGGCGGCGGCGG 1440
Db 1381 CCGCCAACTTTGGGTTCGGGCGGCGCTCGGTTCGTTGTCGGTTCGGCGGCGGCGGCGG 1440

Db 541 TGGCCGCGCTCAGATGTGGGACACGCTGGCGAGTGACCTGTTTTCGCGCGCGTGGCGT 600
QY 601 TTCAGTCCGTGCTCTGGGCTCTGACCGGTGGGTCTGTGATAGTTTCGTGCGCGGTCTGA 660
Db 601 TTCAGTCCGTGCTCTGGGCTCTGACCGGTGGGTCTGTGATAGTTTCGTGCGCGGTCTGA 660
QY 661 TGGTGGCGCGCTCTGCGGTGATGTGGCGTGAAGCGTCAACCGCGGGGAGCCGAGC 720
Db 661 TGGTGGCGCGCTCTGCGGTGATGTGGCGTGAAGCGTCAACCGCGGGGAGCCGAGC 720
QY 721 TGACCGCGCGCGAGTCTGGGTTGTCTGGCGGCTACGAGCGCGTATGGGTGACGG 780
Db 721 TGACCGCGCGCGAGTCTGGGTTGTCTGGCGGCTACGAGCGCGTATGGGTGACGG 780
QY 781 TGCCCGCGCGGTGATCGCCAGAACCGTGTCTGAACCTGATGATCTGATAGCGACCAAC 840
Db 781 TGCCCGCGCGGTGATCGCCAGAACCGTGTCTGAACCTGATGATCTGATAGCGACCAAC 840
QY 841 TCTTGGGCAAAACACCGCGGATCGCGGTCAACGAGCGCGAATACGGCGAGATGGG 900
Db 841 TCTTGGGCAAAACACCGCGGATCGCGGTCAACGAGCGCGAATACGGCGAGATGGG 900
QY 901 CCCAAGACCGCGCGATGTTTGGCTACGCGCGGCGACCGCGCGCGACCGCGCGACGT 960
Db 901 CCCAAGACCGCGCGATGTTTGGCTACGCGCGGCGACCGCGCGCGACCGCGCGACGT 960
QY 961 TGCTGCCGTTTCGAGGAGCGCGGAGATGACAGCGCGGTTGGCTCTCTGAGCAGCGCG 1020
Db 961 TGCTGCCGTTTCGAGGAGCGCGGAGATGACAGCGCGGTTGGCTCTCTGAGCAGCGCG 1020
QY 1021 CCGCGGTTCGAGGAGCGCTCCGACACCGCGCGGCGAACCACTGATCAACAAATGTGCCCC 1080
Db 1021 CCGCGGTTCGAGGAGCGCTCCGACACCGCGCGGCGAACCACTGATCAACAAATGTGCCCC 1080
QY 1081 AGGCGTTCACAGTGGCGCGCGACCGCGCGCGACCGCGCGCTTTCGAGTGGGTG 1140
Db 1081 AGGCGTTCACAGTGGCGCGCGACCGCGCGCGACCGCGCGCTTTCGAGTGGGTG 1140
QY 1141 GCCTGTGGAAGACGCTTCGCGCGATCGCTGCGCGATCAGCACATGTTCCATGCGCCA 1200
Db 1141 GCCTGTGGAAGACGCTTCGCGCGATCGCTGCGCGATCAGCACATGTTCCATGCGCCA 1200
QY 1201 ACAACCAATGTCATGACCACTCGGCTGTGTCGATGACCAACACCTTGAGTCCGATGT 1260
Db 1201 ACAACCAATGTCATGACCACTCGGCTGTGTCGATGACCAACACCTTGAGTCCGATGT 1260
QY 1261 TGAAGGCTTTGCTCCG 1320
Db 1261 TGAAGGCTTTGCTCCG 1320
QY 1321 TCCGCGCGATGAGTCTCGTGGCGAGCTCGTGGGTTCTTCGGGTCTGGCGGTGGGTTG 1380
Db 1321 TCCGCGCGATGAGTCTCGTGGCGAGCTCGTGGGTTCTTCGGGTCTGGCGGTGGGTTG 1380
QY 1381 CCGCGACTTGGTTCGCGCGCGCTCGTGGTTCGTTGTCGTCGCGAGCGCTGGCGCG 1440
Db 1381 CCGCGACTTGGTTCGCGCGCGCTCGTGGTTCGTTGTCGTCGCGAGCGCTGGCGCG 1440
QY 1441 CCGCGAACCGAGGAGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
Db 1441 CCGCGAACCGAGGAGTCAACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1500
QY 1501 CCGCGAAAGAGGCGCGCGCGAGATGTCGCGCGGTGCGCGGTGCGCGCGCGCGCGCG 1560
Db 1501 CCGCGAAAGAGGCGCGCGCGAGATGTCGCGCGGTGCGCGGTGCGCGCGCGCGCGCG 1560
QY 1561 GGGCGGTGTTGGCTCAGTGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
Db 1561 GGGCGGTGTTGGCTCAGTGGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1620
QY 1621 ATTCTCCGCGCGCGGATATCGCCCGCGCGCGCTTGTGCGAGGACCGGTTTCGCGACT 1680
Db 1621 ATTCTCCGCGCGCGGATATCGCCCGCGCGCGCTTGTGCGAGGACCGGTTTCGCGACT 1680

QY 1681 TCCCGCGCTGCGCTCGACCGTCCGCGATGTCGCCCAAGTGGGGCCACAGGTGGTCA 1740
Db 1681 TCCCGCGCTGCGCTCGACCGTCCGCGATGTCGCCCAAGTGGGGCCACAGGTGGTCA 1740
QY 1741 ACATCAACCAAACTGGGCTACAAACACCGCTGGGCGCGCGGACCGGATCGTCATCG 1800
Db 1741 ACATCAACCAAACTGGGCTACAAACACCGCTGGGCGCGCGGACCGGATCGTCATCG 1800
QY 1801 ATCCCAACCGTTCGCTGTCACCAACCAACCACTGATCGCGGCGCGCAACCAATG 1860
Db 1801 ATCCCAACCGTTCGCTGTCACCAACCAACCACTGATCGCGGCGCGCAACCAATG 1860
QY 1861 CGTTACAGCTCGGCTCCGCGCAAACTACGCGCTCGATGTGTCGGGTATGACCGCACCC 1920
Db 1861 CGTTACAGCTCGGCTCCGCGCAAACTACGCGCTCGATGTGTCGGGTATGACCGCACCC 1920
QY 1921 AGGATGTCGCGTTCGAGCTGCGCGGTGCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 1980
Db 1921 AGGATGTCGCGTTCGAGCTGCGCGGTGCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 1980
QY 1981 GCGGCGTTCGCGTTCGAGCTGCGCGGTGCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 2040
Db 1981 GCGGCGTTCGCGTTCGAGCTGCGCGGTGCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 2040
QY 2041 CCGCGCTTCGCGTTCGAGCTGCGCGGTGCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 2100
Db 2041 CCGCGCTTCGCGTTCGAGCTGCGCGGTGCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 2100
QY 2101 CCGTTCAGCTTCGCGTTCGAGCTGCGCGGTGCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 2160
Db 2101 CCGTTCAGCTTCGCGTTCGAGCTGCGCGGTGCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 2160
QY 2161 CCGGCTTCGCGTTCGAGCTGCGCGGTGCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 2220
Db 2161 CCGGCTTCGCGTTCGAGCTGCGCGGTGCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 2220
QY 2221 CCGGCTTCGCGTTCGAGCTGCGCGGTGCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 2280
Db 2221 CCGGCTTCGCGTTCGAGCTGCGCGGTGCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 2280
QY 2281 CCGGCTTCGCGTTCGAGCTGCGCGGTGCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 2340
Db 2281 CCGGCTTCGCGTTCGAGCTGCGCGGTGCGCGTTCGCGTTCGCGTTCGCGTTCGCGTTCGCGT 2340

RESULT 6

US-10-369-983-8
; Sequence 8, Application US/10369983
; Publication No. US2003023593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 8
; LENGTH: 2451
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB81F (MTB72F-DPV)
US-10-369-983-8

Query Match 95.6%; Score 2186.8; DB 16; Length 2451;

Best Local Similarity 99.9%; Pred. No. 0;			
Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
QY	39	CATATGTCATCACCATCACCATCACCAGCGCGGTGCGATACCTCAGCTGCCAGAGT	98
Db	1	CATATGTCATCACCATCACCATCACCAGCGCGGTGCGATACCTCAGCTGCCAGAGT	60
QY	99	GCGCAGGATTCGCCATTCGGATCGGCGCAGGCGATGCGGCGCGATCGATCGATCG	158
Db	61	GCGCAGGATTCGCCATTCGGATCGGCGCAGGCGATGCGGCGCGATCGATCGATCG	120
QY	159	GCTGGGGGTACCCACCTTCATATCGGGGCTTACCGCTTCCTGGCTGGGTGTTGTC	218
Db	121	GCTGGGGGTACCCACCTTCATATCGGGGCTTACCGCTTCCTGGCTGGGTGTTGTC	180
QY	219	GACAAACAGGCAACGGCGCAGGATTCAGCGGTGCTCGGAGGCTCCGCGCGCAAGT	278
Db	181	GACAAACAGGCAACGGCGCAGGATTCAGCGGTGCTCGGAGGCTCCGCGCGCAAGT	240
QY	279	CTCGGCATCTCCACCGGCGACGTGATCACCGGTGCGACGGCGCTCCGATCAACTCGGCC	338
Db	241	CTCGGCATCTCCACCGGCGACGTGATCACCGGTGCGACGGCGCTCCGATCAACTCGGCC	300
QY	339	ACGCGATGCGGACCGCTTAACGGGATCATCCGGTGAGCTCATCTCGGTGACCTGG	398
Db	301	ACGCGATGCGGACCGCTTAACGGGATCATCCGGTGAGCTCATCTCGGTGACCTGG	360
QY	399	CAAAACAAAGTCGGCGCGCGCATCAGGGAACGTGACATTCGCCAGGAGACCCCGGCC	458
Db	361	CAAAACAAAGTCGGCGCGCGCATCAGGGAACGTGACATTCGCCAGGAGACCCCGGCC	420
QY	459	GAATTCATGTTGATTCGGGGGCTTACCAACCGGAGATCAATCCCGGAGGATGACGCC	518
Db	421	GAATTCATGTTGATTCGGGGGCTTACCAACCGGAGATCAATCCCGGAGGATGACGCC	480
QY	519	GGCCCGGCTTCGCTCGTGGCGCGGCTCAGATGTGGGACAGCGTGCGGAGTAC	578
Db	481	GGCCCGGCTTCGCTCGTGGCGCGGCTCAGATGTGGGACAGCGTGCGGAGTAC	540
QY	579	CTGTTTCGCGCGCGTTCAGTTCGTTGCTGCGGTGCTGAGCGTGCGGTGCTGG	638
Db	541	CTGTTTCGCGCGCGTTCAGTTCGTTGCTGCGGTGCTGAGCGTGCGGTGCTGG	600
QY	639	ATAGTTTCGCGCGGTTCGATGTTGGCGCGCTCCCGTATCTGCGGTGATGAC	698
Db	601	ATAGTTTCGCGCGGTTCGATGTTGGCGCGCTCCCGTATCTGCGGTGATGAC	660
QY	699	GTCAACCGCGGCGGCGAGCTGACCGCGCGCGGCTCGGCTTCGCGCGGCTTAC	758
Db	661	GTCAACCGCGGCGGCGAGCTGACCGCGCGCGGCTCGGCTTCGCGCGGCTTAC	720
QY	759	GAGACGCGTATGCGTGAACGTCGCGCGCGCGTTCGCGCGAGACCGTTCGACTG	818
Db	721	GAGACGCGTATGCGTGAACGTCGCGCGCGCGTTCGCGCGAGACCGTTCGACTG	780
QY	819	ATGATTCGATAGCGACCAACCTTTGGGGCAAAACACCGCGCGATCGCGGTCAACGAG	878
Db	781	ATGATTCGATAGCGACCAACCTTTGGGGCAAAACACCGCGCGATCGCGGTCAACGAG	840
QY	879	GCCGAATACCGCGAGATTCGCGCGCAAGACCGCGCGCGATTTGCTACCGCGCGG	938
Db	841	GCCGAATACCGCGAGATTCGCGCGCAAGACCGCGCGCGATTTGCTACCGCGCGG	900
QY	939	ACGCGACCGGACCGGAGCTTCGCTCGCTTCGAGGAGCGCGGAGATGACACGCG	998
Db	901	ACGCGACCGGACCGGAGCTTCGCTCGCTTCGAGGAGCGCGGAGATGACACGCG	960
QY	999	GCTGGGCTTCCTCGAGCAGGCGCGCGGTTCGAGAGGCTTCGACACCGCGCGGCAAC	1058
Db	961	GCTGGGCTTCCTCGAGCAGGCGCGCGGTTCGAGAGGCTTCGACACCGCGCGGCAAC	1020
QY	1059	CAGTTGATGAACAATGTCCCGCAGCGCTGCAACAGCTGCGCGCAGCGAGGCGACC	1118

Db	1021	CAGTTGATGAACAATGTCCCGCAGCGCTGCAACAGCTGCGCGCAGCGAGGCGACC	1080
QY	1119	ACGCTTCTTCCAGCTGGGTGGCTGTGGAAGACGGTCTTCGCGCATCGGTCCCGATC	1178
Db	1081	ACGCTTCTTCCAGCTGGGTGGCTGTGGAAGACGGTCTTCGCGCATCGGTCCCGATC	1140
QY	1179	AGCAACATGTGTGATGCGCAACCAACATGTCGATGACCAACTCGGCTGTGTCGATG	1238
Db	1141	AGCAACATGTGTGATGCGCAACCAACATGTCGATGACCAACTCGGCTGTGTCGATG	1200
QY	1239	ACCAACACCTTGAGCTCGATGTTCAAGGGCTTTGCTCCGCGCGCGCGCCACAGCGCGT	1298
Db	1201	ACCAACACCTTGAGCTCGATGTTCAAGGGCTTTGCTCCGCGCGCGCGCCACAGCGCGT	1260
QY	1299	CAAAACCGCGCGCAAAACCGGGTCCGCGCGATGAGCTTCGCTGGGACGCTCGTGGGTCT	1358
Db	1261	CAAAACCGCGCGCAAAACCGGGTCCGCGCGATGAGCTTCGCTGGGACGCTCGTGGGTCT	1320
QY	1359	TCGGGTCTGGCGGTGGGTGGCGCAACTTGGGTTCGGCGGCTCGGTGGTTCGTTG	1418
Db	1321	TCGGGTCTGGCGGTGGGTGGCGCAACTTGGGTTCGGCGGCTCGGTGGTTCGTTG	1380
QY	1419	TCGGTGGCGCAGGCTGGCGCGCGCAACAGGACGTCAACCGCGCGCGCGCGCGT	1478
Db	1381	TCGGTGGCGCAGGCTGGCGCGCGCAACAGGACGTCAACCGCGCGCGCGCGCGT	1440
QY	1479	CCGTGACGACCTGACAGCGCGCGGAGAGGCGCGCGCAGATGCTGGGCGGCGT	1538
Db	1441	CCGTGACGACCTGACAGCGCGCGGAGAGGCGCGCGCAGATGCTGGGCGGCGT	1500
QY	1539	CCGTGGGCGACAGTGGCGCGCGGTGGTGGGTTCAGTGGTTCGTCGTGTTCCG	1598
Db	1501	CCGTGGGCGACAGTGGCGCGCGGTGGTGGGTTCAGTGGTTCGTCGTGTTCCG	1560
QY	1599	CCGGAACCTTATGTGATGCGCATTCCTCGGACGCGCGCATGTCGCCCGCGCGCTT	1658
Db	1561	CCGGAACCTTATGTGATGCGCATTCCTCGGACGCGCGCATGTCGCCCGCGCGCTT	1620
QY	1659	TCGAGGACCGGTTCCCGGCTTCGCCGCTTCGCGCTTCGCGCTTCGCGATGTCGCC	1718
Db	1621	TCGAGGACCGGTTCCCGGCTTCGCCGCTTCGCGCTTCGCGCTTCGCGATGTCGCC	1680
QY	1719	CAAGTGGGCGCACAGGTGGTCAACATCAACACCAAACTGGGCTACAAACCGCGTGGG	1778
Db	1681	CAAGTGGGCGCACAGGTGGTCAACATCAACACCAAACTGGGCTACAAACCGCGTGGG	1740
QY	1779	GCGGACCGGATGCTCATCGATCCCAACGGTTCGTCGTCGTCGACCAACCAACGATG	1838
Db	1741	GCGGACCGGATGCTCATCGATCCCAACGGTTCGTCGTCGTCGACCAACCAACGATG	1800
QY	1839	GCGGCGCACCGACATCAATGCGTTTCAGCGTCGCTCCGCGCAAACTACGGGTTCGAT	1898
Db	1801	GCGGCGCACCGACATCAATGCGTTTCAGCGTCGCTCCGCGCAAACTACGGGTTCGAT	1860
QY	1899	GTGTCGCGTATGACCGCACCGATGTCGCGGTTCGTCGTCGTCGCGGTTCGCGTGGC	1958
Db	1861	GTGTCGCGTATGACCGCACCGATGTCGCGGTTCGTCGTCGTCGCGGTTCGCGTGGC	1920
QY	1959	CTGCGCTCGGCGCGATCGGTGGCGGTTCGCGGTTCGTCGTCGTCGCGGTTCGCGTGGC	2018
Db	1921	CTGCGCTCGGCGCGATCGGTGGCGGTTCGCGGTTCGTCGTCGTCGCGGTTCGCGTGGC	1980
QY	2019	AAACAGCGTGGGCGAGGCGGAAACCGCGGTTCGTCGTCGTCGCGGTTCGCGTGGC	2078
Db	1981	AAACAGCGTGGGCGAGGCGGAAACCGCGGTTCGTCGTCGTCGCGGTTCGCGTGGC	2040
QY	2079	CAAAACGTCAGGCGGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC	2138
Db	2041	CAAAACGTCAGGCGGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC	2100
QY	2139	CAGTTTCGATGCGCGCATCCAGCGCGGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTC	2198
Db	2101	CAGTTTCGATGCGCGCATCCAGCGCGGTTCGTCGTCGTCGTCGTCGTCGTCGTCGTC	2160

Qy	2199	CAGTGTGCGTATGAACACGGCGCGTC	2228
Dd	2161	CAGTGTGCGTATGAACACGGCGCGTC	2190
RESULT 7			
US-10-369-983-7			
; Sequence 7, Application US/10369983			
; Publication No. US2003023593A1			
; GENERAL INFORMATION:			
; APPLICANT: Skeiky, Yasir			
; APPLICANT: Guderian, Jeff			
; APPLICANT: Reed, Steven			
; APPLICANT: Corixa Corporation			
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis			
; FILE REFERENCE: 014058-009081US			
; CURRENT APPLICATION NUMBER: US/10/369,983			
; CURRENT FILING DATE: 2003-02-18			
; PRIOR APPLICATION NUMBER: US 60/357,351			
; PRIOR FILING DATE: 2002-02-15			
; NUMBER OF SEQ ID NOS: 22			
; SOFTWARE: Patencin Ver. 2.1			
; SEQ ID NO 7			
; LENGTH: 2487			
; TYPE: DNA			
; ORGANISM: Artificial Sequence			
; FEATURE:			
; OTHER INFORMATION: Description of Artificial Sequence:fusion protein			
; OTHER INFORMATION: MTB93P (MTB72F-MTI)			
US-10-369-983-7			
Query Match 95.6%; Score 2186.8; DB 16; Length 2487;			
Best Local Similarity 99.9%; Pred. No. 0;			
Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
Qy	39	CATATGCATCACCATCACACGGCGCGTCCGATACTCCAGCTGCCAGGTT	98
Dd	1	CATATGCATCACCATCACACGGCGCGTCCGATACTCCAGCTGCCAGGTT	60
Qy	99	GCGCAGGATTCCGATTCGATCGGCGAGCGATGCGATCGGCGCGAGATCCGATCG	158
Dd	61	GCGCAGGATTCCGATTCGATCGGCGAGCGATGCGATCGGCGCGAGATCCGATCG	120
Qy	159	GGTGCGGCGTCAACCACCGTTCATACCGGCTACCGCTTCCTCGCTGGGTGTTTC	218
Dd	121	GGTGCGGCGTCAACCACCGTTCATACCGGCTTCCTCGCTGGGTGTTTC	180
Qy	219	GACAACACGGCAACGGCGCACAGTCCAACGCGTGTGCGGAGCGTCCGGCGGCAAGT	278
Dd	181	GACAACACGGCAACGGCGCACAGTCCAACGCGTGTGCGGAGCGTCCGGCGGCAAGT	240
Qy	279	CTCGGCATCTCCACGGCGAGTGTACCGCGTTCAGCGCGTCCCGATCAACTCGGCC	338
Dd	241	CTCGGCATCTCCACGGCGAGTGTACCGCGTTCAGCGCGTCCCGATCAACTCGGCC	300
Qy	339	ACCGGATCGCGACCGCTTAACGGGATCATCCCGGATCATCCCGTGAAGTACCTGG	398
Dd	301	ACCGGATCGCGACCGCTTAACGGGATCATCCCGGATCATCCCGTGAAGTACCTGG	360
Qy	399	CAAAACAGTCCGGCGGACCGTACAGGAACTGATGACCGGAGGACCCCGGCC	458
Dd	361	CAAAACAGTCCGGCGGACCGTACAGGAACTGATGACCGGAGGACCCCGGCC	420
Qy	459	GAATTCATGGTGAATTTCCGGCGGTTACACCGGAGATCAACTCCGGAGGATGTACGCC	518
Dd	421	GAATTCATGGTGAATTTCCGGCGGTTACACCGGAGATCAACTCCGGAGGATGTACGCC	480
Qy	519	GCCCCGGGTTCGCTCGCTGGTGGCGCGCTCAGATGTGGNACAGCGTGGCGATGAC	578
Dd	481	GCCCCGGGTTCGCTCGCTGGTGGCGCGCTCAGATGTGGNACAGCGTGGCGATGAC	540
Qy	579	CTGTGTTTCGGCGCGTCCGCGTTCAGTCGTTGCTGCGGCTCTGACCGTGGGTCGTG	638

1621	Db	TCGACGAGACGGGTTCCGCGCATTTCCCGGGCGCTGCCCTCTCGACCCGCTCGCGATGGTTCGCC	1680
1719	QY	CAAGTGGGGCCACAGGTGGTCAACATCAACACCAAACTGGGCTACAAACACGCGCTGGGC	1778
1681	Db	CAAGTGGGGCCACAGGTGGTCAACATCAACACCAAACTGGGCTACAAACACGCGCTGGGC	1740
1779	QY	GCGGGACCGGCATCGTCATCGATCCACAGGTGCTGCTGACCAACACACACACGTGATC	1838
1741	Db	GCGGGACCGGCATCGTCATCGATCCACAGGTGCTGCTGACCAACACACACGTGATC	1800
1839	QY	GCGGGGCCACCGACATCAATGGGTTTCAGCGTCGGGTCGCGCCAAACCTACGCGTCGAT	1898
1801	Db	GCGGGGCCACCGACATCAATGGGTTTCAGCGTCGGGTCGCGCCAAACCTACGCGTCGAT	1860
1899	QY	GTGGTCGGGTATGACCGCACCCACAGGATGTCGCGGTCTGCAGCTGCGCGGTGCCGTGGC	1958
1861	Db	GTGGTCGGGTATGACCGCACCCACAGGATGTCGCGGTCTGCAGCTGCGCGGTGCCGTGGC	1920
1959	QY	CTCCGCTCGCGCGCGATCGGTGCGCGGTGTCGCGTTGGTAGCCCGTCGTCGCGATGGGC	2018
1921	Db	CTCCGCTCGCGCGCGATCGGTGCGCGGTGTCGCGTTGGTAGCCCGTCGTCGCGATGGGC	1980
2019	QY	AAACAGCGGTGGGACGGCGGAAACGCCCGTCGCGGTGCTGCGAGGGTGGTTCGCGCTCGGC	2078
1981	Db	AAACAGCGGTGGGACGGCGGAAACGCCCGTCGCGGTGCTGCGAGGGTGGTTCGCGCTCGGC	2040
2079	QY	CAAAACCGTGCAGCGCTCGGATTCGCTGACCGGTGTCGGAAGAGACATTGAAACGGTTGATC	2138
2041	Db	CAAAACCGTGCAGCGCTCGGATTCGCTGACCGGTGTCGGAAGAGACATTGAAACGGTTGATC	2100
2139	QY	CAGTTTCGATCCCGCGATCCAGCCGGTGATTTCGGGGGGGGCCCGTCGTCACAGCCCTAGGA	2198
2101	Db	CAGTTTCGATCCCGCGATCCAGCCGGTGATTTCGGGGGGGGCCCGTCGTCACAGCCCTAGGA	2160
2199	QY	CAGGTGGTTCGGTATGAACACGGCGCGGTCC	2228
2161	Db	CAGGTGGTTCGGTATGAACACGGCGCGGTCC	2190

RESULT 8

```

US-10-369-983-6
; Sequence 6, Application US/10369983
; Publication No. US2003023593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 2637
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:fusion protein
; OTHER INFORMATION: MTB89F (MTB72F-Erd14)
; US-10-369-983-6

```

Query Match	95.6%	Score 2186.8	DB 16	Length 2637
Best Local Similarity	99.9%	Pred. No. 0		
Matches 2188	Conservative 0	Mismatches 2	Indels 0	Gaps 0
Qy	39	CATATGATCACCATCACCATCACACGGCGCGTCCGATAACTTCCAGCTGTCCACGGT	98	
Db	1	CATATGATCACCATCACCATCACACGGCGCGTCCGATAACTTCCAGCTGTCCACGGT	60	

Qy	99	GGCAGGGATTCCCATTCGGATCGGCGAGGGCATGGCGATCGCGGGCCAGATCCGATCG	158
Db	61	GGCAGGGATTCCCATTCGGATCGGCGAGGGCATGGCGATCGCGGGCCAGATCCGATCG	120
Qy	159	GGTGGGGGTCAACCACCGTTTCAATTCGGGCGCTACCGCCTTCTCGGCTTGGGTGTGTC	218
Db	121	GGTGGGGGTCAACCACCGTTTCAATTCGGGCGCTACCGCCTTCTCGGCTTGGGTGTGTC	180
Qy	219	GACAAACAACGGCAACGGCGCAAGAGTCAACCGGTGGTCGGGAGCGCTCCGGCGGCAAGT	278
Db	181	GACAAACAACGGCAACGGCGCAAGAGTCAACCGGTGGTCGGGAGCGCTCCGGCGGCAAGT	240
Qy	279	CTCGGATCTCCACCGGCGACGTGATCACGGCGTCCGACGGCGCTCCGATCAACTCGGCC	338
Db	241	CTCGGATCTCCACCGGCGACGTGATCACGGCGTCCGACGGCGCTCCGATCAACTCGGCC	300
Qy	339	ACCGGATCGGCAACCGCTTAAACGGGCAATATCCCGGTGAACGTATCTCGGTGACCTGG	398
Db	301	ACCGGATCGGCAACCGCTTAAACGGGCAATATCCCGGTGAACGTATCTCGGTGACCTGG	360
Qy	399	CAAAACAAGTCGGCGGCGACGGTACAGGAGAGTGACATTCGCGAGGAGACCCCGGCC	458
Db	361	CAAAACAAGTCGGCGGCGACGGTACAGGAGAGTGACATTCGCGAGGAGACCCCGGCC	420
Qy	459	GAATTATGGTGATTTTCGGGCGTTTACCACCGGAGATCAACTCCCGAGGATGTACGCC	518
Db	421	GAATTATGGTGATTTTCGGGCGTTTACCACCGGAGATCAACTCCCGAGGATGTACGCC	480
Qy	519	GGCCCGGGTTGGCCCTCGCTGTGGCGCGGCTCAGATGTGGGACAGCGTCGCGAGTGAC	578
Db	481	GGCCCGGGTTGGCCCTCGCTGTGGCGCGGCTCAGATGTGGGACAGCGTCGCGAGTGAC	540
Qy	579	CTGTTTTCGGCGCGCTCGGGTTTCAAGTCGGTGGTCTGGGGTCTGACGGTGGGGTCTGG	638
Db	541	CTGTTTTCGGCGCGCTCGGGTTTCAAGTCGGTGGTCTGGGGTCTGACGGTGGGGTCTGG	600
Qy	639	ATAGGTTGTCGGCGGTCTGATGTGGCGCGGCTCGCGTATGTGGCTGGATGAGC	698
Db	601	ATAGGTTGTCGGCGGTCTGATGTGGCGCGGCTCGCGTATGTGGCTGGATGAGC	660
Qy	699	GTACACGGGGCAGCCGAGCTGACCGCGCCAGTCCGGTTCCTCGCGCGGCTAC	758
Db	661	GTACACGGGGCAGCCGAGCTGACCGCGCCAGTTCGGTTCCTCGCGCGGCTAC	720
Qy	759	GAGACGGGTATGGGTGACGFTGCCCGCCCGGTGATCGCGAGAACCGTGTGAACTG	818
Db	721	GAGACGGGTATGGGTGACGFTGCCCGCCCGGTGATCGCGAGAACCGTGTGAACTG	780
Qy	819	ATGATTTGATAGGACCAACTCTTTGGGGCAAAACACCCCGCGATCGCGGTCAACGAG	878
Db	781	ATGATTTGATAGGACCAACTCTTTGGGGCAAAACACCCCGCGATCGCGGTCAACGAG	840
Qy	879	GCCGAAATACCGCGAGATGTGGGCCCAAGACCGCGCCCGCATGTTTGGCTACGCCGGGG	938
Db	841	GCCGAAATACCGCGAGATGTGGGCCCAAGACCGCGCCCGCATGTTTGGCTACGCCGGGG	900
Qy	939	ACGGCGACGGGACGGGAGCTTGTCTCGGTCGAGGAGGCGCGGAGATGACCAAGCGG	998
Db	901	ACGGCGACGGGACGGGAGCTTGTCTCGGTCGAGGAGGCGCGGAGATGACCAAGCGG	960
Qy	999	GGTGGGCTCTCTCAGACAGGCGCGCGGTTCGAGGAGGCTCCGACACCGCGCGGCGAAC	1058
Db	961	GGTGGGCTCTCTCAGACAGGCGCGCGGTTCGAGGAGGCTCCGACACCGCGCGGCGAAC	1020
Qy	1059	CAGTTGATGAACAATGTGCCCAAGGCGCTGCAACAGCTGGCCAGCCCAACGCAAGGAC	1118
Db	1021	CAGTTGATGAACAATGTGCCCAAGGCGCTGCAACAGCTGGCCAGCCCAACGCAAGGAC	1080
Qy	1119	ACGCTTCTTCAAGCTGGGTGCTGTGGAAGACGGTCTCGCCGATCGGTTCGCGATC	1178
Db	1081	ACGCTTCTTCAAGCTGGGTGCTGTGGAAGACGGTCTCGCCGATCGGTTCGCGATC	1140
Qy	1179	AGCAACATGTGTGATGGCGCAACCAACATGTGATGATCAACATCGGGTGTGTGATG	1238

```
1141 AGCAACATGGTGTGATGGCCAAACACACATGTCGATGACCAACTCGGTGTGCGATG 1200
1239 ACCAACACCTTGAAGTGTGAAGGCTTTGCTCCGGCGCGCGCCGCCAGGCGGTG 1298
1201 ACCAACACCTTGAAGTGTGAAGGCTTTGCTCCGGCGCGCGCCGCCAGGCGGTG 1260
1299 CAATCCGCGCGCAAAACGGGTCGCGCGGATGAGCTCGTGGGCGAGCTCGCTGGGTTCT 1358
1261 CAATCCGCGCGCAAAACGGGTCGCGCGGATGAGCTCGTGGGCGAGCTCGCTGGGTTCT 1320
1359 TCGGGTCTGGCGGTGGGGTGGCGGCAACTTGGGTCGGCGCGCTCGGTTCGTTG 1418
1321 TCGGGTCTGGCGGTGGGGTGGCGGCAACTTGGGTCGGCGCGCTCGGTTCGTTG 1380
1419 TCGGTCGCGGAGGCTGGCGCGGCGCAACACGAGCAGTCAACCGCGCGCGCGCGCGCTG 1478
1381 TCGGTCGCGGAGGCTGGCGCGGCGCAACACGAGCAGTCAACCGCGCGCGCGCGCGCTG 1440
1479 CCGCTGACACGCTGACCGCGCGCGGAAAGAGGCGCGGCGAGATGCTGGGCGGGCTG 1538
1441 CCGCTGACACGCTGACCGCGCGCGGAAAGAGGCGCGGCGAGATGCTGGGCGGGCTG 1500
1539 CCGTGGGCGAGATGGCGCGCGAGGCGGCTGGGTCAGTGGTGTGCTGCTGCTTCG 1598
1501 CCGTGGGCGAGATGGCGCGCGAGGCGGCTGGTGGGCTCAGTGGTGTGCTGCTTCG 1560
1599 CCGGACCTTATGATGCGCGATCTCCGCGAGCGCGGATATCGCCCGCGCGCTTG 1658
1561 CCGGACCTTATGATGCGCGATCTCCGCGAGCGCGGATATCGCCCGCGCGCTTG 1620
1659 TCGCAGGACCGGTTGCGCGATCTCCGCGCGCTGCGCTCGACCCGCTCGCGATGCTG 1718
1621 TCGCAGGACCGGTTGCGCGATCTCCGCGCGCTGCGCTCGACCCGCTCGCGATGCTG 1680
1719 CAAGTGGGCGCAGAGTGTCAATCAACACCAAACTGGGCTACACACGCGGCTGGC 1778
1681 CAAGTGGGCGCAGAGTGTCAATCAACACCAAACTGGGCTACACACGCGGCTGGC 1740
1779 GCCGGGACCGGCTATCGATCGATCCCAACGCTGTGCTGTGACCAACCAACGCTGATC 1838
1741 GCCGGGACCGGCTATCGATCGATCCCAACGCTGTGCTGTGACCAACCAACGCTGATC 1800
1839 GCCGGGCGCAGCAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTG 1898
1801 GCCGGGCGCAGCAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTGAGTGTGCTG 1860
1899 GTGTCGGGTATGACCGGATGACCGGATGACCGGATGACCGGATGACCGGATGACCGG 1958
1861 GTGTCGGGTATGACCGGATGACCGGATGACCGGATGACCGGATGACCGGATGACCGG 1920
1959 CTCGCGTGGCGGCGATCGGTGGCGGCTGCGGCTGGTGGTGGTGGTGGTGGTGGTGG 2018
1921 CTCGCGTGGCGGCGATCGGTGGCGGCTGCGGCTGGTGGTGGTGGTGGTGGTGGTGG 1980
2019 AACAGCGTGGCGAGGCGGAAACCGCGGTCGCTGCGGAGGCTGGTGGGCTCGGCTCG 2078
1981 AACAGCGTGGCGAGGCGGAAACCGCGGTCGCTGCGGAGGCTGGTGGGCTCGGCTCG 2040
2079 CAATCCGTCAGGCGTGGGATTCGCTGACCGGTCGCGGAGAGACATGAAACGGTTGATC 2138
2041 CAATCCGTCAGGCGTGGGATTCGCTGACCGGTCGCGGAGAGACATGAAACGGTTGATC 2100
2139 CAGTTCCATCGCGATGACCGCGGTTGATTCGCGGCGGCGGCTGCTGCAACGCGCTAGGA 2198
2101 CAGTTCCATCGCGATGACCGCGGTTGATTCGCGGCGGCGGCTGCTGCAACGCGCTAGGA 2160
2199 CAGTGGTTCGGTATGAACACGCGCGGCTCC 2228
2161 CAGTGGTTCGGTATGAACACGCGCGGCTCC 2190
```

```
US-10-098-732A-64
; Sequence 64, Application US/10098732A
; Publication NO. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 64
; LENGTH: 2808
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB72F-MAPS
; OTHER INFORMATION: (795F) fusion construct, TB MTB72F (Ra12-FbH9-Ra35)
; OTHER INFORMATION: linked to Leishmania thiol-specific antioxidant
; OTHER INFORMATION: (TSA or MAPS)
US-10-098-732A-64
```

```
Query Match 95.6%; Score 2186.8; DB 15; Length 2808;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 39 CATATGATCATCACCATCACATCACACGCGCGCTCCGATTAATTCAGCTGTCCAGGGT 98
DB 1 CATATGATCATCACCATCACATCACACGCGCGCTCCGATTAATTCAGCTGTCCAGGGT 60
QY 99 GGGCAGGATTTCGGATTCGGATCGGCGAGCGATCGGCGGCGAGATCCGATCG 158
DB 61 GGGCAGGATTTCGGATTCGGATCGGCGAGCGATCGGCGGCGAGATCCGATCG 120
QY 159 GGTGGGGGTCAACCAACCGTTTCATATCGGCGCTACCGCTTCCTCGGCTTGGGTGTGTC 218
DB 121 GGTGGGGGTCAACCAACCGTTTCATATCGGCGCTACCGCTTCCTCGGCTTGGGTGTGTC 180
QY 219 GACAAACAGCGACAGCGGCGACGAGTCAACGCGTGTTCGGAGCGCTCCGGCGGCAAGT 278
DB 181 GACAAACAGCGACAGCGGCGACGAGTCAACGCGTGTTCGGAGCGCTCCGGCGGCAAGT 240
QY 279 CTCGGCATCTCCACCGCGAGCTGATCACCGCGGTTCAGCGCGCTCCGATCAACTCGGCC 338
DB 241 CTCGGCATCTCCACCGCGAGCTGATCACCGCGGTTCAGCGCGCTCCGATCAACTCGGCC 300
QY 339 ACCCGATTCGGCGAGCGCTTAACGGGATATCCCGGTGACGTCATCTCGGTGACCTGG 398
DB 301 ACCCGATTCGGCGAGCGCTTAACGGGATATCCCGGTGACGTCATCTCGGTGACCTGG 360
QY 399 CAATCCAGTTCGGCGGCGACGCGTACAGGAACTGACATTTGGCGGAGGACCCCGGCC 458
DB 361 CAATCCAGTTCGGCGGCGACGCGTACAGGAACTGACATTTGGCGGAGGACCCCGGCC 420
QY 459 GAATTCATGTTGGATTTTCGGGGCGTTTACCACCGGAGATCAACTCCCGAGGATGTACGCC 518
DB 421 GAATTCATGTTGGATTTTCGGGGCGTTTACCACCGGAGATCAACTCCCGAGGATGTACGCC 480
QY 519 GCGCGGGTTCGGCTCGCTGGTGGCGCGCTCAGATGTGGGACAGCTGGCGAGTGCAC 578
DB 481 GCGCGGGTTCGGCTCGCTGGTGGCGCGCTCAGATGTGGGACAGCTGGCGAGTGCAC 540
QY 579 CTGTTTTTCGGCGCGCTCGGCGTTTCAGTTCGGTGTGCTGGGCTCTGACGGTGGGTCTGG 638
DB 541 CTGTTTTTCGGCGCGCTCGGCGTTTCAGTTCGGTGTGCTGGGCTCTGACGGTGGGTCTGG 600
QY 639 ATAGGTTTCGGCGCGGTTCGATGTTGGTGGCGGCGCTCCGCCGTATGTGGCGTGTAGCG 698
```

Db 601 ATAGTTCTGCGGGTCTGATGTTGGCGCGGCTCCCGTATGTGGCTGATGAGC 660
QY 699 GTACCGCGGACAGCCGAGTGTACCGCGCCCAAGTCCCGGTTCGTGCGGCGCCCTAC 758
Db 661 GTACCGCGGACAGCCGAGTGTACCGCGCCCAAGTCCCGGTTCGTGCGGCGCCCTAC 720
QY 759 GAGACGGCTATGGCTGACGTTGCGCGCGCGGCTGATCGCGGAGAACCGTGTCAACTG 818
Db 721 GAGACGGCTATGGCTGACGTTGCGCGCGCGGCTGATCGCGGAGAACCGTGTCAACTG 780
QY 819 ATGATTCTGATAGCGACCAACCTCTTTGGGGCAAAACACCCCGCGCATCGCGTCAACGAG 878
Db 781 ATGATTCTGATAGCGACCAACCTCTTTGGGGCAAAACACCCCGCGCATCGCGTCAACGAG 840
QY 879 GCGGATACGGAGATGTGGGCGCCAGAGCGCGCGCGATGTTGGCTACGCCCGGGG 938
Db 841 GCGGATACGGAGATGTGGGCGCCAGAGCGCGCGCGATGTTGGCTACGCCCGGGG 900
QY 939 AGGGCGACGGCGACGGCGACGTTGCTGCGGTTTCGAGAGGCGCGGAGATACCAAGCGCG 998
Db 901 AGGGCGACGGCGACGGCGACGTTGCTGCGGTTTCGAGAGGCGCGGAGATGACAGCGGG 960
QY 999 GGTGGGCTCTCGAGCAGCGCGCGGCTCGAGAGGCTTCGACACCGCGCGCGCGAAC 1059
Db 961 GGTGGGCTCTCGAGCAGCGCGCGGCTCGAGAGGCTTCGACACCGCGCGCGCGAAC 1020
QY 1059 CAGTTGATCAACATGTGCGCCAGCGCTGCAACAGCTGCGCCAGCCAGCGAGGCGACC 1118
Db 1021 CAGTTGATCAACATGTGCGCCAGCGCTGCAACAGCTGCGCCAGCCAGCGAGGCGACC 1080
QY 1119 AGCGCTCTTTCAAGCTGGGTGGCTGTGGAGAGCGTCTCGCCGATCGGTTCGCCGATC 1178
Db 1081 AGCGCTCTTTCAAGCTGGGTGGCTGTGGAGAGCGTCTCGCCGATCGGTTCGCCGATC 1140
QY 1179 AGCAACATGGTGTGATGTCGACCAACACACATGTGATGACCACTCGGGTGTGCGATG 1238
Db 1141 AGCAACATGGTGTGATGTCGACCAACACACATGTGATGACCACTCGGGTGTGCGATG 1200
QY 1239 ACCAACACCTTGAGCTCGATGTTGAAGGGCTTTGCTCCGCGCGCGCGCGCGCGCGGTG 1298
Db 1201 ACCAACACCTTGAGCTCGATGTTGAAGGGCTTTGCTCCGCGCGCGCGCGCGCGGTG 1260
QY 1299 CAAACCGCGCGCAAAACCGGGTTCGCGGCGATGAGCTCGCTGGGAGCTCGTGGGTCTT 1358
Db 1261 CAAACCGCGCGCAAAACCGGGTTCGCGGCGATGAGCTCGCTGGGAGCTCGTGGGTCTT 1320
QY 1359 TCGGGTCTGGGCGGTGGGTGGCGCAACTTGGGTGGGCGCGCGCTCGGTGCGGTTCGTTG 1418
Db 1321 TCGGGTCTGGGCGGTGGGTGGCGCAACTTGGGTGGGCGCGCGCTCGGTGCGGTTCGTTG 1380
QY 1419 TCGGTGCGCGAGCTGGGCGCGCGCAACCGAGGAGTCAACCGCGCGCGCGCGCGGTG 1478
Db 1381 TCGGTGCGCGAGCTGGGCGCGCGCAACCGAGGAGTCAACCGCGCGCGCGCGCGGTG 1440
QY 1479 CCGCTACCACTGTACCGCGCGCGCGCAAGAGGCGCGCGAGATGCTGGGCGGGTGTG 1538
Db 1441 CCGCTACCACTGTACCGCGCGCGCGCAAGAGGCGCGCGAGATGCTGGGCGGGTGTG 1500
QY 1539 CCGGTGGGCGAGATGGGCGCGCGAGGCGCGGTGTTGGGCTCAGTGGTGTGCTGGGTTCG 1598
Db 1501 CCGGTGGGCGAGATGGGCGCGCGAGGCGCGGTGTTGGGCTCAGTGGTGTGCTGGGTTCG 1560
QY 1599 CCGGACCTATGTATGCGGCACTTCGCGGAGCGCGGCGATATCGCCCGCGCGCGCTTG 1658
Db 1561 CCGGACCTATGTATGCGGCACTTCGCGGAGCGCGGCGATATCGCCCGCGCGCGCTTG 1620
QY 1659 TCGCAGGACCGGTTCGCGCACTTCGCGGCGCTGCGCGCTCGACCGCTTCGCGATGTCGCG 1718
Db 1621 TCGCAGGACCGGTTCGCGCACTTCGCGGCGCTGCGCGCTCGACCGCTTCGCGATGTCGCG 1680
QY 1719 CAAAGTGGGCGACAGTGTGTCAACATCAACAACTGGGTGTACAAACCGCGCGGTGGGC 1778

Db 1681 CAAGTGGGCGCACAGGTGGTCAACATCAACACCAACTGGGCTACAAACCGCGGTGGGC 1740
QY 1779 GCGGGAACGGGATCGTATCGATCCCAACCGGTGTCGTGTCGACCAACACCGGTGATC 1838
Db 1741 GCGGGAACGGGATCGTATCGATCCCAACCGGTGTCGTGTCGACCAACACCGGTGATC 1800
QY 1839 GCGGGAACGGGATCGTATCGATCCCAACCGGTGTCGTGTCGACCAACCTACGGGCTGAT 1898
Db 1801 GCGGGAACGGGATCGTATCGATCCCAACCGGTGTCGTGTCGACCAACCTACGGGCTGAT 1860
QY 1899 GTGTCGGGTATGACCCACCGAGATGTCGGGTGTCGAGTGCAGTGCAGGTCGGGTGGC 1958
Db 1861 GTGTCGGGTATGACCCACCGAGATGTCGGGTGTCGAGTGCAGTGCAGGTCGGGTGGC 1920
QY 1959 CTGCGGTGCGCGCGATCGGTGCGCGGTGTCGGGTGTCGAGTGCAGTGCAGGTCGGGTGGC 2018
Db 1921 CTGCGGTGCGCGCGATCGGTGCGCGGTGTCGGGTGTCGAGTGCAGTGCAGGTCGGGTGGC 1980
QY 2019 AACAGCGGTGGGACGGGCGGAAACCGCGGTGTCGGGTGTCGAGGTCGGGTGGCCTCGGC 2078
Db 1981 AACAGCGGTGGGACGGGCGGAAACCGCGGTGTCGGGTGTCGAGGTCGGGTGGCCTCGGC 2040
QY 2079 CAACCGGTGAGGCGGTGCGATTCGTCGACCGGTGCGGAGAGACATTGAACGGGTTCATC 2138
Db 2041 CAACCGGTGAGGCGGTGCGATTCGTCGACCGGTGCGGAGAGACATTGAACGGGTTCATC 2100
QY 2139 CAGTTCGATGCGCGGATCCAGCGCGGTGATTTCGGGCGCGCGCGGTTCGTCGTCGTCGAG 2198
Db 2101 CAGTTCGATGCGCGGATCCAGCGCGGTGATTTCGGGCGCGCGCGGTTCGTCGTCGTCGAG 2160
QY 2199 CAGTTCGATGCGCGGATCCAGCGCGGTGATTTCGGGCGCGCGCGGTTCGTCGTCGTCGAG 2228
Db 2161 CAGTTCGATGCGCGGATCCAGCGCGGTGATTTCGGGCGCGCGCGGTTCGTCGTCGTCGAG 2190

RESULT 10

US-10-369-983-5

; Sequence 5, Application US/10369983

; Publication No. US20030235593A1

; GENERAL INFORMATION:

; APPLICANT: Skeiky, Yasir

; APPLICANT: Guderian, Jeff

; APPLICANT: Reed, Steven

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis

; FILE REFERENCE: 014059-009081US

; CURRENT APPLICATION NUMBER: US/10/369,983

; PRIOR FILING DATE: 2003-02-18

; PRIOR APPLICATION NUMBER: US 60/357,351

; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 5

; LENGTH: 2808

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: fusion protein

; OTHER INFORMATION: R95F (MTB72F-MAPS)

US-10-369-983-5

Query Match 95.6%; Score 2186.8; DB 16; Length 2808;

Best Local Similarity 99.9%; Pred. No. 0;

Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 39 CATATGATCATCAATCAACCATCAACCGGCGCGGTTCGATTAACCTTCAGCTGTGCCAGGGT 98

Db 1 CATATGATCATCAATCAACCATCAACCGGCGCGGTTCGATTAACCTTCAGCTGTGCCAGGGT 60

QY 99 GGGCAGGATTCGCATTCGATCGGCGAGCGATGCGATCGCGGCGCGAGATCCGATCG 158

Db 61 GGGCAGGATTCGCATTCGATCGGCGAGCGATGCGATCGCGGCGCGAGATCCGATCG 120

QY 159 GGTGGGGGTCACCACCGTTTATATCGGGCTTACCGCTTCTCGCTTGGGTGTGTC 218
DB 121 GGTGGGGGTCACCACCGTTTATATCGGGCTTACCGCTTCTCGCTTGGGTGTGTC 180
QY 219 GACAAACGCGCAACGCGGACAGGTCCAGCGGTGTGCGGAGCGCTCGGGGCAAGT 278
DB 181 GACAAACGCGCAACGCGGACAGGTCCAGCGGTGTGCGGAGCGCTCGGGGCAAGT 240
QY 279 CTCGGCATCTCACCGGCGAGTGTATCACCGCGGTGACGCGCTCCGATCAACTCGGCC 338
DB 241 CTCGGCATCTCACCGGCGAGTGTATCACCGCGGTGACGCGCTCCGATCAACTCGGCC 300
QY 339 ACCCGGATGGGAGCGGCTTAAACGGGATCATCCCGGTGACGTCTCTCGGTGACCTGG 398
DB 301 ACCCGGATGGGAGCGGCTTAAACGGGATCATCCCGGTGACGTCTCTCGGTGACCTGG 360
QY 399 CAACACCAAGTCGGGCGGACGCGTACAGGGAAACGTGACATTTGGCCGAGGACCCCGGCC 458
DB 361 CAACACCAAGTCGGGCGGACGCGTACAGGGAAACGTGACATTTGGCCGAGGACCCCGGCC 420
QY 459 GAATTCATGTGTGATTTTCGGGGCGTTTACCAACCGAGATCAACTCCCGGAGGATGTAGGCC 518
DB 421 GAATTCATGTGTGATTTTCGGGGCGTTTACCAACCGAGATCAACTCCCGGAGGATGTAGGCC 480
QY 519 GCGCGGGTTCGGGCTCGCTGCTGTGGCGCGGCTCAGATGTGGACAGCGTGGCGAGTGAC 578
DB 481 GCGCGGGTTCGGGCTCGCTGCTGTGGCGCGGCTCAGATGTGGACAGCGTGGCGAGTGAC 540
QY 579 CTGTTTTCGGCGGCTCGGCGTTTCACTCGGTGTCTGGGCTCTGACGCTGGGGTCTGTG 638
DB 541 CTGTTTTCGGCGGCTCGGCGTTTCACTCGGTGTCTGGGCTCTGACGCTGGGGTCTGTG 600
QY 639 ATAGGTTCTCGGGGTCTGATGTGGCGCGGCTCTCGGATGTGGCGTGTGATGTGGCGT 698
DB 601 ATAGGTTCTCGGGGTCTGATGTGGCGCGGCTCTCGGATGTGGCGTGTGATGTGGCGT 660
QY 699 GTACCGCGGGGAGGCGAGTGTACCGCGCGCGCGGCTCGGGTGTGCGCGGCGGCTTAC 758
DB 661 GTACCGCGGGGAGGCGAGTGTACCGCGCGCGCGGCTCGGGTGTGCGCGGCGGCTTAC 720
QY 759 GAGACGGGTATGGGTGACGGTGTGCGCGCGCGGCTGATCGCGAGAACCGTGTGTAAGT 818
DB 721 GAGACGGGTATGGGTGACGGTGTGCGCGCGCGGCTGATCGCGAGAACCGTGTGTAAGT 780
QY 819 ATGATTCTGATAGGACCAACCTCTTGGGGCAAAACACCCCGCGGATCGGGTCAACGAG 878
DB 781 ATGATTCTGATAGGACCAACCTCTTGGGGCAAAACACCCCGCGGATCGGGTCAACGAG 840
QY 879 GCCGAATACGGCGAGATGTGGGCGCAAGACCGCGCGCGATTTGGCTACGCGCGGG 938
DB 841 GCCGAATACGGCGAGATGTGGGCGCAAGACCGCGCGCGATTTGGCTACGCGCGGG 900
QY 939 ACGCGGACGGCGAGCGGCTTGTCTCGGTTCGAGGAGCGCGGAGATGACCGCGG 998
DB 901 ACGCGGACGGCGAGCGGCTTGTCTCGGTTCGAGGAGCGCGGAGATGACCGCGG 960
QY 999 GGTGGGTCTCTCGAGGCGCGCGGCTCGAGGAGCGCTTCGACACCGCGCGGCGAAC 1058
DB 961 GGTGGGTCTCTCGAGGCGCGCGGCTCGAGGAGCGCTTCGACACCGCGCGGCGAAC 1020
QY 1059 CAGTTGATGACATGTGCGCGAGGCGCTGACAGCTGCGCGCGCGCGAGCGGACCG 1118
DB 1021 CAGTTGATGACATGTGCGCGAGGCGCTGACAGCTGCGCGCGCGCGAGCGGACCG 1080
QY 1119 ACGCTTCTTCCAGCTGGGTGGCTGTGGAAGACGCTCTCGCGGATCGGTGCGCGATC 1178
DB 1081 ACGCTTCTTCCAGCTGGGTGGCTGTGGAAGACGCTCTCGCGGATCGGTGCGCGATC 1140
QY 1179 AGCAATGTGTGATGGCGCAACCAACCATGTGATGACCAACTCGGGTGTGTGATG 1238
DB 1141 AGCAATGTGTGATGGCGCAACCAACCATGTGATGACCAACTCGGGTGTGTGATG 1200
QY 1239 ACCAACACCTTGAGTCTGATGTTGAAGGGCTTGTGCTCGCGCGCGCGCGCGCGCTG 1298

RESULT 11

US-10-369-983-11

; Sequence, 11, Application US/10369983

; Publication No. US2003023559A1

; GENERAL INFORMATION:

DB 1201 ACCAACACCTTGATCTGATGTTGAAGGGCTTGTCTCGGGCGGGCGCGCCAGCGGTG 1260
QY 1299 CAAACCGGGCGCAAAACGGGTTCGGGCGGATGAGTCTGGGCGAGTCTCGCTGGGTCT 1358
DB 1261 CAAACCGGGCGCAAAACGGGTTCGGGCGGATGAGTCTGGGCGAGTCTCGCTGGGTCT 1320
QY 1359 TCGGTCTTGGGCGGTGGGTTCGGCCAACTTGGGTTCGGGCGGCTCGGTTCGGTTCGTTG 1418
DB 1321 TCGGTCTTGGGCGGTGGGTTCGGCCAACTTGGGTTCGGGCGGCTCGGTTCGGTTCGTTG 1380
QY 1419 TCGGTTCGGCGAGCTTCGGGCGCGCCAAACAGGAGTCAACCGCGCGCGCGCGCGGTG 1478
DB 1381 TCGGTTCGGCGAGCTTCGGGCGCGCCAAACAGGAGTCAACCGCGCGCGCGCGCGGTG 1440
QY 1479 CCGCTGACAGCTGACAGCGCGCGGAAAGAGGGCCCGGCGAGATGCTGGGCGGGCTG 1538
DB 1441 CCGCTGACAGCTGACAGCGCGCGGAAAGAGGGCCCGGCGAGATGCTGGGCGGGCTG 1500
QY 1539 CCGGTGGGCGAGTGGGCGCGCGAGGCGCGGTGGTGGGCTCAGTGTGTGCTGCGGTTCGG 1598
DB 1501 CCGGTGGGCGAGTGGGCGCGCGAGGCGCGGTGGTGGGCTCAGTGTGTGCTGCGGTTCGG 1560
QY 1599 CCGGACCCCTATGTGATGTCGCGCATTTCCGCGAGCGCGGCGATATCGCCCGCGGCTTG 1658
DB 1561 CCGGACCCCTATGTGATGTCGCGCATTTCCGCGAGCGCGGCGATATCGCCCGCGGCTTG 1620
QY 1659 TCGGAGGACCGGTTCGCGACTTCCCGCGGTGCCCTTCGACCCGTCGCGGATGTCGCC 1718
DB 1621 TCGGAGGACCGGTTCGCGACTTCCCGCGGTGCCCTTCGACCCGTCGCGGATGTCGCC 1680
QY 1719 CAAAGTGGGCGCACAGGTGTCAACATCAACACCAAACTGGGCTACAAACACCGCGTGGG 1778
DB 1681 CAAAGTGGGCGCACAGGTGTCAACATCAACACCAAACTGGGCTACAAACACCGCGTGGG 1740
QY 1779 CCGGAGACCGGATCGTTCATCGATCCCAACCGGTGTGCTGTGACAAACACCGTGTGTC 1838
DB 1741 CCGGAGACCGGATCGTTCATCGATCCCAACCGGTGTGCTGTGACAAACACCGTGTGTC 1800
QY 1839 CCGGCGCGACCGGATCAATCGCTTACGCTTCGGCTCCGCGCAAACTACGGGTCGAT 1898
DB 1801 CCGGCGCGACCGGATCAATCGCTTACGCTTCGGCTCCGCGCAAACTACGGGTCGAT 1860
QY 1899 GTGCTCGGCTATGACCGCACCCAGGATGTCCGGTGTGCTGACAGTGTGCGGCTGCGGTCG 1958
DB 1861 GTGCTCGGCTATGACCGCACCCAGGATGTCCGGTGTGCTGACAGTGTGCGGTCGCGTGGC 1920
QY 1959 CTGCGCTCGGCGGATCGGTGGGCGGTTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2018
DB 1921 CTGCGCTCGGCGGATCGGTGGGCGGTTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1980
QY 2019 AACAGCGGTGGCGAGGCGGAAACCGCGTGGCTTCGGAGGGTGGTTCGGCTTCGGC 2078
DB 1981 AACAGCGGTGGCGAGGCGGAAACCGCGTGGCTTCGGAGGGTGGTTCGGCTTCGGC 2040
QY 2079 CAAACCGTGCAGGCGGTTCGATTCGCTGACCGGTTCGCGAGAGACATTTGAACCGGTTGATC 2138
DB 2041 CAAACCGTGCAGGCGGTTCGATTCGCTGACCGGTTCGCGAGAGACATTTGAACCGGTTGATC 2100
QY 2139 CAGTTGATGCGCGATCCAGCCCGGTGATTCGGGCGGGCGCGCTCGTCAACGGGCTTAGGA 2198
DB 2101 CAGTTGATGCGCGATCCAGCCCGGTGATTCGGGCGGGCGCGCTCGTCAACGGGCTTAGGA 2160
QY 2199 CAGTGTGCTGATGAACACCGCGCGTTC 2228
DB 2161 CAGTGTGCTGATGAACACCGCGCGTTC 2190

APPLICANT: Skeiky, Yasir
APPLICANT: Guderian, Jeff
APPLICANT: Reed, Steven
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
FILE REFERENCE: 014058-009081US
CURRENT APPLICATION NUMBER: US/10/369,983
CURRENT FILING DATE: 2003-02-18
PRIOR APPLICATION NUMBER: US 60/357,351
PRIOR FILING DATE: 2002-02-15
NUMBER OF SEQ ID NOS: 22
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 11
LENGTH: 3060
TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: fusion protein
OTHER INFORMATION: MTB103F (MTB72F-85b)

US-10-369-983-11

Query Match 95.6%; Score 2186.8; DB 16; Length 3060;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	39	CATATGCATCACCATCACCATCACACGGCGCGTCCGATAACTTCACGCTGTCCAGGCT	98
DB	1	CATATGCATCACCATCACCATCACACGGCGCGTCCGATAACTTCACGCTGTCCAGGCT	60
QY	99	GGCGAGGATTCGCCATTCCGATTCGGGACGGGATGGGATCGCGGCGCAGATCCGATCG	158
DB	61	GGCGAGGATTCGCCATTCCGATTCGGGACGGGATGGGATCGCGGCGCAGATCCGATCG	120
QY	159	GGTGGGGGTCAACCAAGTTCATATCGGGCTACCGGCTTCTCGGCTGGGTGTGTC	218
DB	121	GGTGGGGGTCAACCAAGTTCATATCGGGCTTACCGGCTTCTCGGCTGGGTGTGTC	180
QY	219	GACAAACAGCGCAACCGGCGACAGTCCAAACGCGTGGTCCGGAGCGCTCCGGCGCAAGT	278
DB	181	GACAAACAGCGCAACCGGCGACAGTCCAAACGCGTGGTCCGGAGCGCTCCGGCGCAAGT	240
QY	279	CTCGGATCTTCAACCGGCGACGTATCACCGGCTGACCGGCTCCGATCAACTCGGCC	338
DB	241	CTCGGATCTTCAACCGGCGACGTATCACCGGCTGACCGGCTCCGATCAACTCGGCC	300
QY	339	ACCGGATGGCGACGCGCTTAACGGGCTATCCCGGTGAGTCACTTCGGTGACCTGG	398
DB	301	ACCGGATGGCGACGCGCTTAACGGGCTATCCCGGTGAGTCACTTCGGTGACCTGG	360
QY	399	CAAACCAAGTCGGCGGCGACGGTACAGGAAAGTGAATGGCCAGGGAACCGCGGCC	458
DB	361	CAAACCAAGTCGGCGGCGACGGTACAGGAAAGTGAATGGCCAGGGAACCGCGGCC	420
QY	459	GAATTCATGGTATTCGGGGGTTACACCGGAGATCAATCCCGGAGGATGACGCC	518
DB	421	GAATTCATGGTATTCGGGGGTTACACCGGAGATCAATCCCGGAGGATGACGCC	480
QY	519	GGCCCGGGTTCGGCTTCGTGGTGGCGGGGTTCAGATGGGACAGCGTGGCGAGTGAC	578
DB	481	GGCCCGGGTTCGGCTTCGTGGTGGCGGGGTTCAGATGGGACAGCGTGGCGAGTGAC	540
QY	579	CTGTTTTCGGCGCGCTTCGGGTTTCACTGGTGGTTCGGGTTGACGGTGGGTTGCG	638
DB	541	CTGTTTTCGGCGCGCTTCGGGTTTCACTGGTGGTTCGGGTTGACGGTGGGTTGCG	600
QY	639	ATAGGTTTCGGCGGGTTCGATGGTGGGGCGGCTCCCGGTATGTGGCTGGATGAGC	698
DB	601	ATAGGTTTCGGCGGGTTCGATGGTGGGGCGGCTCCCGGTATGTGGCTGGATGAGC	660
QY	699	GTACCGCGGGGACGCGGAGTGAACCGCGCCAGGTCGGGTTCTCGCGCGGCTTAC	758
DB	661	GTACCGCGGGGACGCGGAGTGAACCGCGCCAGGTCGGGTTCTCGCGCGGCTTAC	720

QY	759	GAGACGGCTATGGCTGACGGTCCCGCGGTGATCCCGAGAACCTGCTGAAC	818
DB	721	GAGACGGCTATGGCTGACGGTCCCGCGGTGATCCCGAGAACCTGCTGAAC	780
QY	819	ATGATTTCTGATAGCGACCAACCTCTTTGGGCAAAAACACCCCGCGATCGCGGTCAACG	878
DB	781	ATGATTTCTGATAGCGACCAACCTCTTTGGGCAAAAACACCCCGCGATCGCGGTCAACG	840
QY	879	GCCGATACGGCGAGATGTGGGCCCAAGACCGCGCGCGATGTTGCTACGCGCGCGG	938
DB	841	GCCGATACGGCGAGATGTGGGCCCAAGACCGCGCGCGATGTTGCTACGCGCGCGG	900
QY	939	ACGGCGACGGCGACGGCGAGTTCCTGCTCCCTTCGAGGAGCGCGCGAGATGACACAGCGCG	998
DB	901	ACGGCGACGGCGACGGCGAGTTCCTGCTCCCTTCGAGGAGCGCGCGAGATGACACAGCGCG	960
QY	999	GGTGGGTCTCTGAGCAGCGCGCGCGTTCGAGAGGCTTCGACACCGCGCGCGCGAAC	1058
DB	961	GGTGGGTCTCTGAGCAGCGCGCGCGTTCGAGAGGCTTCGACACCGCGCGCGCGAAC	1020
QY	1059	CAGTTGATGACATGTGCCCGCGCTGCAACAGCTGGCCAGCCACCGCAGGCGAC	1118
DB	1021	CAGTTGATGACATGTGCCCGCGCTGCAACAGCTGGCCAGCCACCGCAGGCGAC	1080
QY	1119	ACGCTTTCTTCCAAAGTGGGTGCTCTGGAAGACGGTCTCGCGCGATCGCTCGCGATC	1178
DB	1081	ACGCTTTCTTCCAAAGTGGGTGCTCTGGAAGACGGTCTCGCGCGATCGCTCGCGATC	1140
QY	1179	AGCAACATGTGTGATGGCGCAACACCATGTCGATGACCACTCGGGTGTGTCGATG	1238
DB	1141	AGCAACATGTGTGATGGCGCAACACCATGTCGATGACCACTCGGGTGTGTCGATG	1200
QY	1239	ACCAACACCTTGAAGTGGGTGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1298
DB	1201	ACCAACACCTTGAAGTGGGTGCTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	1260
QY	1299	CAAAACCGCGCGCAAAACCGGGTCCGGCGGATGAGTCTCGTGGGACGCTCGCTGGGTCT	1358
DB	1261	CAAAACCGCGCGCAAAACCGGGTCCGGCGGATGAGTCTCGTGGGACGCTCGCTGGGTCT	1320
QY	1359	TCGGGTCTGGGCGGTGGGCGCGCAACTTGGGTTCGGCGCGGCTCGGTTCGTTG	1418
DB	1321	TCGGGTCTGGGCGGTGGGCGCGCAACTTGGGTTCGGCGCGGCTCGGTTCGTTG	1380
QY	1419	TCGGTGGCGCAGGCTGGGCGCGCAACAGCAGTCAACCGCGCGCGCGCGCGCGCG	1478
DB	1381	TCGGTGGCGCAGGCTGGGCGCGCAACAGCAGTCAACCGCGCGCGCGCGCGCGCG	1440
QY	1479	CCGCTGACGACCTGACAGCGCGCGGAGAGGCGCGCGCGAGATGCTGGGCGGCGTG	1538
DB	1441	CCGCTGACGACCTGACAGCGCGCGGAGAGGCGCGCGCGAGATGCTGGGCGGCGTG	1500
QY	1539	CCGCTGGGCGCAGTGGGCGCGCGCGGTGGGTGCTCAGTGGTGTGCTGCTGCTTCCG	1598
DB	1501	CCGCTGGGCGCAGTGGGCGCGCGCGGTGGGTGCTCAGTGGTGTGCTGCTGCTTCCG	1560
QY	1599	CCGCGACCTTATGTGATGCGCATTCCTCGGCGACCGCGGATTCGCGCGCGCGCGCTG	1658
DB	1561	CCGCGACCTTATGTGATGCGCATTCCTCGGCGACCGCGGATTCGCGCGCGCGCGCTG	1620
QY	1659	TCGAGAGACCGGTTTCGCGCACTTCCTCGCGCTGCTCCCTCGACCGCTCGCGATGCTGC	1718
DB	1621	TCGAGAGACCGGTTTCGCGCACTTCCTCGCGCTGCTCCCTCGACCGCTCGCGATGCTGC	1680
QY	1719	CAAGTGGGCGCACAGGTGTCACATCAACACCAAAATGGGCTACAAACACCGCGTGGC	1778
DB	1681	CAAGTGGGCGCACAGGTGTCACATCAACACCAAAATGGGCTACAAACACCGCGTGGC	1740
QY	1779	GCCTGGGACCGGATCGATCCCAACGGTCTGCTGCTGACCAACCAACACCGTGCATC	1838
DB	1741	GCCTGGGACCGGATCGATCCCAACGGTCTGCTGCTGACCAACCAACACCGTGCATC	1800
QY	1839	CGCGCGCGCACCGACATCAATGCTGCTTACGCGTGGGCTCCGCGCGCAACCTACGCGCTCGAT	1898

Db 1801 GCGGGCGCCCGGACATCAATGCGTTTACGGTTCGGCTCCGCCAAACCTACGGCGTCGAT 1860
Qy 1899 GTGGTCCGGTATGACCGCACCCAGGATGTCCGGTGTCTGAGCTGCGCGGTGCGGGTGC 1958
Db 1861 GTGGTCCGGTATGACCGCACCCAGGATGTCCGGTGTCTGAGCTGCGCGGTGCGGGTGC 1920
Qy 1959 CTCGGCTCCGGCGGATCGGTGGCGGGTGGCGGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 2018
Db 1921 CTGGCTCCGGCGGATCGGTGGCGGGTGGCGGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1980
Qy 2019 AACAGCGGTGGCGAGGCGGAAACCCCGCTGCGGTGCTGCGAGGGTGGTGGTGGTGGTGG 2078
Db 1981 AACAGCGGTGGCGAGGCGGAAACCCCGCTGCGGTGCTGCGAGGGTGGTGGTGGTGGTGG 2040
Qy 2079 CAACCGGTGGCGAGGCGGATTCGCTGACCGGTGGCGAGGATTCGCTGACCGGTGGTGG 2138
Db 2041 CAACCGGTGGCGAGGCGGATTCGCTGACCGGTGGCGAGGATTCGCTGACCGGTGGTGG 2100
Qy 2139 CAGTTTCGATGCCGATCCAGCCCGGTGATTCGGGCGGGCGCGTGGTGGTGGTGGTGGTGG 2198
Db 2101 CAGTTTCGATGCCGATCCAGCCCGGTGATTCGGGCGGGCGCGTGGTGGTGGTGGTGG 2160
Qy 2199 CAGTGGTTCGATGAACACGCGCGCTCC 2228
Db 2161 CAGTGGTTCGATGAACACGCGCGCTCC 2190

RESULT 12
US-10-369-983-10
; Sequence 10, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 10
; LENGTH: 3104
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB102tm2F (MTB102FTM, MTB72F-HTCC#1)
US-10-369-983-10

Query Match 95.6%; Score 2186.8; DB 16; Length 3104;
Best Local Similarity 99.9%; Pred. No. 0; Mismatches 2; Indels 0; Gaps 0;
Matches 2186; Conservative 0

Qy 39 CATATGATCACCATACCATACACGCGCGGTCCGATAACTTCCAGCTGTCCAGGTT 98
Db 1 CATATGATCACCATACCATACACGCGCGGTCCGATAACTTCCAGCTGTCCAGGTT 60
Qy 99 GGGCAGGGATTCGCATTCGATCGGCGAGCGATCGGATCGGCGGCGAGATCCGATCG 158
Db 61 GGGCAGGGATTCGCATTCGATCGGCGAGCGATCGGATCGGCGGCGAGATCCGATCG 120
Qy 159 GGTGGGGGTTCACCCACCGTTTCATATCGGCGCTACCGCTTCCTCGGCTTGGGTGTTC 218
Db 121 GGTGGGGGTTCACCCACCGTTTCATATCGGCGCTACCGCTTCCTCGGCTTGGGTGTTC 180
Qy 219 GACACACCGGCAACGCGGACGATCCACGCTGCTGGGAGCGCTCCGGGCGCAAGT 278
Db 181 GACACACCGGCAACGCGGACGATCCACGCTGCTGGGAGCGCTCCGGGCGCAAGT 240

Qy 279 CTCGGCATCTCACCAGGCGAGTGTATCACCAGGCTGACGCGGCTCCGATCAACTCGGCC 338
Db 241 CTCGGCATCTCACCAGGCGAGTGTATCACCAGGCTGACGCGGCTCCGATCAACTCGGCC 300
Qy 339 ACCGCGATGGCGGACCGCTTAAACGGGATCATCCCGGTGACGTCATCTCGGTGACCTGG 398
Db 301 ACCGCGATGGCGGACCGCTTAAACGGGATCATCCCGGTGACGTCATCTCGGTGACCTGG 360
Qy 399 CAACCAAGTCCGGCGGCGACCGGTACAGGGAACGTGACATTTGGCCGAGAGGACCCCGGCC 458
Db 361 CAACCAAGTCCGGCGGCGACCGGTACAGGGAACGTGACATTTGGCCGAGAGGACCCCGGCC 420
Qy 459 GAATTCATGTTGGATTTTCGGGCGGTTACCAACCGGAGATCAACTCCGAGGATGTCAGCC 518
Db 421 GAATTCATGTTGGATTTTCGGGCGGTTACCAACCGGAGATCAACTCCGAGGATGTCAGCC 480
Qy 519 GGGCCGGGTTCCGGCTCGCTCGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTGAC 578
Db 481 GGGCCGGGTTCCGGCTCGCTCGTGGCGCGGCTCAGATGTGGGACAGCGTGGCGAGTGAC 540
Qy 579 CTGTTTTCCGCGCGGTTCGGCGGTTTTCAGTCCGTTGCTGGGGTCTGACGTTGGGTCTGG 638
Db 541 CTGTTTTCCGCGCGGTTCGGCGGTTTTCAGTCCGTTGCTGGGGTCTGACGTTGGGTCTGG 600
Qy 639 ATAGTTTCGTCGGCGGGTCTGATGTTGGCGGCGGCTCCGCGTATGTGGCGTGGATGAGC 698
Db 601 ATAGTTTCGTCGGCGGGTCTGATGTTGGCGGCGGCTCCGCGTATGTGGCGTGGATGAGC 660
Qy 699 GTACCCGGGGGAGCGCGGACTGACCGCGCCCGCAGGTCGCGGTTGCTCGCGGCGCTAC 758
Db 661 GTACCCGGGGGAGCGCGGACTGACCGCGCCCGCAGGTCGCGGTTGCTCGCGGCGCTAC 720
Qy 759 GAGACGGGTATGGCTGACCGTGCCTCCCGCGGTGATCGCGGAGAACCGTCTGAACTG 818
Db 721 GAGACGGGTATGGCTGACCGTGCCTCCCGCGGTGATCGCGGAGAACCGTCTGAACTG 780
Qy 819 ATGATTCGATAGCGAACCACTCTTTGGGGCAAAACACCCCGCGGATCGCGGTCAAAG 878
Db 781 ATGATTCGATAGCGAACCACTCTTTGGGGCAAAACACCCCGCGGATCGCGGTCAAAG 840
Qy 879 GCGGAATACGCGGAGATGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 938
Db 841 GCGGAATACGCGGAGATGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900
Qy 939 ACGGCGACGGCGACGGCGACCTGCTGCGGTTTCGAGGAGGCGCGCGGAGATGACAGCGCG 998
Db 901 ACGGCGACGGCGACGGCGACCTGCTGCGGTTTCGAGGAGGCGCGCGGAGATGACAGCGCG 960
Qy 999 GGTGGGCTTCGAGCAGGCGCGCGGTCGAGAGGCGCTCCGACACGCGCGCGCGGCGAAC 1058
Db 961 GGTGGGCTTCGAGCAGGCGCGCGGTCGAGAGGCGCTCCGACACGCGCGCGCGGCGAAC 1020
Qy 1059 CAGTTGATGAACAATGTGCCCCAGGCGCTGCAACAGGTGGCCAGCCAGCCAGCGGCGACC 1118
Db 1021 CAGTTGATGAACAATGTGCCCCAGGCGCTGCAACAGGTGGCCAGCCAGCGGCGACC 1080
Qy 1119 ACGCCCTTCCTCAAGCTGGGTGGCTGTGGAAGACGGTCTCGCGCGATCGGTTCGCCGATC 1178
Db 1081 ACGCCCTTCCTCAAGCTGGGTGGCTGTGGAAGACGGTCTCGCGCGATCGGTTCGCCGATC 1140
Qy 1179 AGCAACATGTTGATGGCCAAACCAACATGTCGATGACCAACTCGGTTGTCGATG 1238
Db 1141 AGCAACATGTTGATGGCCAAACCAACATGTCGATGACCAACTCGGTTGTCGATG 1200
Qy 1239 ACCAACACTTGAAGCTGATGTTGAAGGCTTTGCTCCGGCGCGGCGCGCGGCGGCGG 1298
Db 1201 ACCAACACTTGAAGCTGATGTTGAAGGCTTTTCTCCGGCGCGGCGCGGCGGCGGCGG 1260
Qy 1299 CAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1358
Db 1261 CAACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320

```
QY 1359 TCGGCTCTGGGCGGTGGCGCCCAACTTGGGTTCGGGTTCGGGCGGCGCTCGGTGGTTCGTTG 1418
Db 1321 TCGGCTCTGGGCGGTGGCGCCCAACTTGGGTTCGGGTTCGGGCGGCGCTCGGTGGTTCGTTG 1380
QY 1419 TCGGCTCCCGCAGCGCTGGCGCGCGGCGCAACACAGGAGAGTCAACCCCGCGCGCGCGCGCGT 1478
Db 1381 TCGGCTCCCGCAGCGCTGGCGCGCGGCGCAACACAGGAGAGTCAACCCCGCGCGCGCGCGCGT 1440
QY 1479 CGGCTACACAGCTGACCAAGCGCGCGGCGCAAGAGGCGCGCGCGAGATGTCGGCGCGGCTG 1538
Db 1441 CGGCTACACAGCTGACCAAGCGCGCGGCGCAAGAGGCGCGCGCGAGATGTCGGCGCGGCTG 1500
QY 1539 CGGCTCGGCGAGATGGCGCGCGGCGCGGCGCTGAGTGGGCTCAGTGGTGTGCTGGTGTTCG 1598
Db 1501 CGGCTCGGCGAGATGGCGCGCGGCGCGGCGCTGAGTGGTGTGCTGGTGTTCG 1560
QY 1599 CGCGACCTATGTATGCGGATTCCTCGGAGCGCGCGGAGATATGCGCCCGCGCGCGCTTG 1658
Db 1561 CGCGACCTATGTATGCGGATTCCTCGGAGCGCGCGGAGATATGCGCCCGCGCGCGCTTG 1620
QY 1659 TCGCAGGACCGGTTCCGCGACATTCGCGCGCTGCGCGCTGCGCGCTGCGCGATGTCGCG 1718
Db 1621 TCGCAGGACCGGTTCCGCGACATTCGCGCGCTGCGCGCTGCGCGCTGCGCGATGTCGCG 1680
QY 1719 CAAGTGGGCGCCACAGGTGGTCAACATCAACACCAACTGGGCTACAAACACCGCGTGGGC 1778
Db 1681 CAAGTGGGCGCCACAGGTGGTCAACATCAACACCAACTGGGCTACAAACACCGCGTGGGC 1740
QY 1779 GCGGACCGGAGATGTCATCGATCCCAAGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1838
Db 1741 GCGGACCGGAGATGTCATCGATCCCAAGGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800
QY 1839 GCGGCGCGCACCGACATCAATGGTTCAGCGTTCGCTGCGCGCTGCGCGCTGCGCTGCGAT 1898
Db 1801 GCGGCGCGCACCGACATCAATGGTTCAGCGTTCGCTGCGCGCTGCGCGCTGCGCTGCGAT 1860
QY 1899 GTGGTTCGGTATGACCGCACCGAGATGTCGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1958
Db 1861 GTGGTTCGGTATGACCGCACCGAGATGTCGGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1920
QY 1959 CTCGCTCGCGCGAGATCGGTGGCGCGCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2018
Db 1921 CTCGCTCGCGCGAGATCGGTGGCGCGCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1980
QY 2019 AACAGCGGTGGGCGAGGCGGCAACCGCGCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2078
Db 1981 AACAGCGGTGGGCGAGGCGGCAACCGCGCTGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
QY 2079 CAACCGTGCAGCGCTCGATTCGCTGACCGGTGCGCGAGAGACATTAACCGGTGCTGCTGCTGCT 2138
Db 2041 CAACCGTGCAGCGCTCGATTCGCTGACCGGTGCGCGAGAGACATTAACCGGTGCTGCTGCTGCTGCT 2100
QY 2139 CAGTTCGATCCCGGATCCAGCGCGGTGATTCGGGCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 2198
Db 2101 CAGTTCGATCCCGGATCCAGCGCGGTGATTCGGGCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCT 2160
QY 2199 CAGGTGGTTCGGTATGAACAGCGCGCGGTCC 2228
Db 2161 CAGGTGGTTCGGTATGAACAGCGCGCGGTCC 2190
```

RESULT 13
US-10-369-983-9
; Sequence 9, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderman, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983

; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 3474
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB114F (MTB72F-mTCC#2)
US-10-369-983-9

Query Match 95.6%; Score 2186.8; DB 16; Length 3474;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2188; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 39 CATATGATCACCATCAACCATCAACCGCGCGCTTCCATCGGCGCTTCCATCGGCTTGGGTGTTGTC 218
Db 1 CATATGATCACCATCAACCATCAACCGCGCGCTTCCATCGGCGCTTCCATCGGCTTGGGTGTTGTC 60
QY 99 GGGCAGGAGTTCGCATTCGATCGGCGAGGAGTGGGATCGGCGCGAGATCGGCGCGAGATCGGATCG 158
Db 61 GGGCAGGAGTTCGCATTCGATCGGCGAGGAGTGGGATCGGCGCGAGATCGGCGCGAGATCGGATCG 120
QY 159 GGTGGGCGGTCAACCAACCGCTTCCATCGGCGCTTCCATCGGCGCTTCCATCGGCTTGGGTGTTGTC 218
Db 121 GGTGGGCGGTCAACCAACCGCTTCCATCGGCGCTTCCATCGGCGCTTCCATCGGCTTGGGTGTTGTC 180
QY 219 GACAAACAGGCAACCGGCGAGTCCAGCGGTGTCGAGCGCTTCCGCGCGCTTCCGATCAATCGGCG 278
Db 181 GACAAACAGGCAACCGGCGAGTCCAGCGGTGTCGAGCGGTGTCGAGCGCTTCCGATCAATCGGCG 240
QY 279 CTCGCGATCTCCACCGCGAGTGCATCACCGCGGTTCAGCGCGCTTCCGATCAATCGGCGCTTCCGATCAATCGGCG 338
Db 241 CTCGCGATCTCCACCGCGAGTGCATCACCGCGGTTCAGCGCGCTTCCGATCAATCGGCGCTTCCGATCAATCGGCG 300
QY 339 ACCCGGATGCGCGAGCGCTTAAACCGGCGATCAATCCCGGTGACGCTATCTCGGTGACGCTG 398
Db 301 ACCCGGATGCGCGAGCGCTTAAACCGGCGATCAATCCCGGTGACGCTATCTCGGTGACGCTG 360
QY 399 CAACCAAGTTCGGGCGCGAGCGGTACAGGAGACGTCATTCGCGCGAGGAGACCGCGCGCG 458
Db 361 CAACCAAGTTCGGGCGCGAGCGGTACAGGAGACGTCATTCGCGCGAGGAGACCGCGCGCG 420
QY 459 GAATTCATGTTGGATTTTCGGGCGGTTCACCGCGAGATCAACTCCGCGAGGATGATGCGCC 518
Db 421 GAATTCATGTTGGATTTTCGGGCGGTTCACCGCGAGATCAACTCCGCGAGGATGATGCGCC 480
QY 519 GCGCGCGGTTCGCGCTTCGTTGGCGCGCGCTCAGATGCGGAGACGAGTGGCGAGTGCAC 578
Db 481 GCGCGCGGTTCGCGCTTCGTTGGCGCGCGCTCAGATGCGGAGACGAGTGGCGAGTGCAC 540
QY 579 CTGTTTCGCGCGGTTCGCGCTTCAGTTCGTTGGGTCTGAGCGGTGCGGTTCGTTG 638
Db 541 CTGTTTCGCGCGGTTCGCGCTTCAGTTCGTTGGGTCTGAGCGGTGCGGTTCGTTG 600
QY 639 ATAGTTCGTCGCGCGGTTCGATGCGTGGCGCGCGCTTCGCGTATGCGGTGCGATGAGC 698
Db 601 ATAGTTCGTCGCGCGGTTCGATGCGTGGCGCGCGCTTCGCGTATGCGGTGCGATGAGC 660
QY 699 GTCACCGCGCGGCGAGCGCGAGTGCACCGCGCGCGAGTTCGCGGTTCGTCGCGCGCGCTTAC 758
Db 661 GTCACCGCGCGGCGAGCGCGAGTGCACCGCGCGCGAGTTCGCGGTTCGTCGCGCGCGCTTAC 720
QY 759 GAGACCGGTATGGGCTGACCGGTTCGCGCGCGGTGATTCGCGGAGAACCGGTGCTGCTGCTGCTGCT 818
Db 721 GAGACCGGTATGGGCTGACCGGTTCGCGCGCGGTGATTCGCGGAGAACCGGTGCTGCTGCTGCTGCT 780
QY 819 ATGATTCTGATAGCGACCAACTCTTGGGCGGAGAAACACCGCGCGATCGCGGTCAACGAG 878
Db 781 ATGATTCTGATAGCGACCAACTCTTGGGCGGAGAAACACCGCGCGATCGCGGTCAACGAG 840

QY 879 CCCGATACGGGAGATGTGGCCCAAGACGCCGCCCGCGATGTTGGCTACGGCGGGCG 938
DB 841 CCCGAATACGGCGAGATGTGGCCCAAGACGCCGCCCGCGATGTTGGCTACGGCGGGCG 900
QY 939 ACGGCGACGGGAGCGGACGTTGCTGCGGTTCCAGGAGGCGCGGAGATGACACAGCGG 998
DB 901 ACGGCGACGGGAGCGGCGGCTGCTGCGGTTCCAGGAGGCGCGGAGATGACACAGCGG 960
QY 999 GGTGGGCTCTCGAGCAGGCGCGCGCGGCTCGAGAGGCGCTCCGACACCGCGCGGCGAAC 1058
DB 961 GGTGGGCTCTCGAGCAGGCGCGCGCGGCTCGAGAGGCGCTCCGACACCGCGCGGCGAAC 1020
QY 1059 CAGTTGATGAACAATGTGCCCGACGCGCTGCAACAGCTGCGCCAGCCACGCAAGGCGACC 1118
DB 1021 CAGTTGATGAACAATGTGCCCGACGCGCTGCAACAGCTGCGCCAGCCACGCAAGGCGACC 1080
QY 1119 ACGGCTTTCTTCCAGCTGGGTGGGCTGTGGAAGACGCTCTCGCGGCAATCGGTCGCGATC 1178
DB 1081 ACGGCTTTCTTCCAGCTGGGTGGGCTGTGGAAGACGCTCTCGCGGCAATCGGTCGCGATC 1140
QY 1179 AGCAACATGGTTCGATGCGCCCAACACACATGTCGATGACCACTCGGCTGTGCGATG 1238
DB 1141 AGCAACATGGTTCGATGCGCCCAACACACATGTCGATGACCACTCGGCTGTGCGATG 1200
QY 1239 ACCAACACCTTGAGCTCGATGTTGAAGGGCTTTGCTCCGGCGGCGGCCGCCAGGCGGTG 1298
DB 1201 ACCAACACCTTGAGCTCGATGTTGAAGGGCTTTGCTCCGGCGGCGGCCGCCAGGCGGTG 1260
QY 1299 CAACACCGGGCGGCAAAACGGGGTCCGGCGGATGAGCTCGCTGGGCGAGCTCGCTGGGTTCT 1358
DB 1261 CAACACCGGGCGGCAAAACGGGGTCCGGCGGATGAGCTCGCTGGGCGAGCTCGCTGGGTTCT 1320
QY 1359 TCGGGTCTGGGCGGTGGGGTGGCGGCAACTTGGGTCCGGCGGCTCGGTCGGTTCGTTG 1418
DB 1321 TCGGGTCTGGGCGGTGGGGTGGCGGCAACTTGGGTCCGGCGGCTCGGTCGGTTCGTTG 1380
QY 1419 TCGGTGCGGCAAGGCTGGGCGGCGGCAACACAGCAGTCAACCGCGGCGGCGGCGGCTG 1478
DB 1381 TCGGTGCGGCAAGGCTGGGCGGCGGCAACACAGCAGTCAACCGCGGCGGCGGCGGCTG 1440
QY 1479 CCGCTGACGAGCTGACCGCGCGGCAAGAGGCGCGGCGGCGAGTCTGCTGGCGGCGGCTG 1538
DB 1441 CCGCTGACGAGCTGACCGCGCGGCAAGAGGCGCGGCGGCGAGTCTGCTGGCGGCGGCTG 1500
QY 1539 CCGGTGGGCGAGATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 1598
DB 1501 CCGGTGGGCGAGATGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTG 1560
QY 1599 CCGGACCTTATGATGCGGCAATTCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCTG 1658
DB 1561 CCGGACCTTATGATGCGGCAATTCGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCTG 1620
QY 1659 TCGCAGGACCGGTTTCGCGGCACTTCCCGCGGCTGCGGCTCGACCCGCTCGGCGAGTGGCGC 1718
DB 1621 TCGCAGGACCGGTTTCGCGGCACTTCCCGCGGCTGCGGCTCGACCCGCTCGGCGAGTGGCGC 1680
QY 1719 CAAGTGGGCGCACAGTGGTCAACATCAACACAACTGGGCTTACAAACACCGCGTGGCG 1778
DB 1681 CAAGTGGGCGCACAGTGGTCAACATCAACACAACTGGGCTTACAAACACCGCGTGGCG 1740
QY 1779 GCCGGACCGGCGATCGTCATCGATCCCAACGCTGCTGCTGACCAACCAACACGCTGATC 1838
DB 1741 GCCGGACCGGCGATCGTCATCGATCCCAACGCTGCTGCTGACCAACCAACACGCTGATC 1800
QY 1839 GCGGCGCCACCGACATCAATGCTTCAGCGTGGGCTTCGCGGCGGCGGCGGCGGCTGAT 1898
DB 1801 GCGGCGCCACCGACATCAATGCTTCAGCGTGGGCTTCGCGGCGGCGGCGGCGGCTGAT 1860
QY 1899 GTGGTGGGCTATGACCGCACCGGAGTGTGCGGCTGCTGAGCTGCGGCTGCGGCTGCGG 1958
DB 1861 GTGGTGGGCTATGACCGCACCGGAGTGTGCGGCTGCTGAGCTGCGGCTGCGGCTGCGG 1920

QY 1959 CTGCGTGGGCGGCGATCGGTGGGCGGCGTGGGTTGGTGGAGCCGCTCGCTCGCGATGGGC 2018
DB 1921 CTGCGTGGGCGGCGATCGGTGGGCGGCGTGGGTTGGTGGAGCCGCTCGTGGCGATGGGC 1980
QY 2019 AACACGCGTGGGCGAGGCGGAAACGCGCGGCGGCGTGGGTTGGTGGAGCGGCTGGGC 2078
DB 1981 AACACGCGTGGGCGAGGCGGAAACGCGCGGCGGCGTGGGTTGGTGGAGCGGCTGGGC 2040
QY 2079 CHAACCGTGGGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2138
DB 2041 CHAACCGTGGGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2100
QY 2139 CAGTTGCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2198
DB 2101 CAGTTGCGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2160
QY 2199 CAGGTGGTGGGTATGAACACACGCGCGCGTCC 2228
DB 2161 CAGGTGGTGGGTATGAACACACGCGCGCGTCC 2190

RESULT 14

US-09-886-349A-17
; Sequence 17, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 2190
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB72MutSA
; OTHER INFORMATION: (Ra12-TbHp-Ra35MutSA) cDNA
US-09-886-349A-17

Query Match 95.5%; Score 2185.2; DB 12; Length 2190;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2187; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 42 ATGATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATC 101
DB 1 ATGATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATC 60
QY 102 CAGGATTCGCATTCGCATTCGCATTCGCATTCGCATTCGCATTCGCATTCGCATTCGCATTC 161
DB 61 CAGGATTCGCATTCGCATTCGCATTCGCATTCGCATTCGCATTCGCATTCGCATTCGCATTC 120
QY 162 GGGGGTCAACCCACCGTTTCATATCGGCGCTACCGCTTCCTCGGCTTGGGTGGTGGTGGTGG 221
DB 121 GGGGGTCAACCCACCGTTTCATATCGGCGCTACCGCTTCCTCGGCTTGGGTGGTGGTGGTGG 180
QY 222 AACAAACGCAACGCGCAACGAGTCCAAACGCGTGGTTCGGAGCGCTTCGGCGGCAAGTCTC 281
DB 181 AACAAACGCAACGCGCAACGAGTCCAAACGCGTGGTTCGGAGCGCTTCGGCGGCAAGTCTC 240
QY 282 GGCATCTCCACCGCGAGCTGATCAGCGCGTTCGACGCGCTCCGATCAACTCGGCCACC 341
DB 241 GGCATCTCCACCGCGAGCTGATCAGCGCGTTCGACGCGCTCCGATCAACTCGGCCACC 300


```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 2190
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72FmutSA
; OTHER INFORMATION: (Ra12-TbH9-Ra35MutSA)
US-10-098-732A-17

Query Match      95.5%; Score 2185.2; DB 15; Length 2190;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2187; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 42 ATGCATCACCATCACCATCACCACGCGCGCGTCCGATAACTTCCAGCTGTCCAGGGTGGG 101
DB 1 ATGCATCACCATCACCATCACCACGCGCGCGTCCGATAACTTCCAGCTGTCCAGGGTGGG 60

QY 102 CAGGGATTCGCATTCGATCGGCGAGCGGATGCGGATCGGCGGCGGATCGGATCGGGT 161
DB 61 CAGGGATTCGCATTCGATCGGCGAGCGGATGCGGATCGGCGGCGGATCGGATCGGGT 120

QY 162 GGGGGTCAACACCGTTCATATCGGCGCTACCGCTTCTCGGCTTGGGTGTTGCGAC 221
DB 121 GGGGGTCAACACCGTTCATATCGGCGCTACCGCTTCTCGGCTTGGGTGTTGCGAC 180

QY 222 AACAAAGCGCAACGCGACGATCCAAACGCGTGGTTCGGGAGCGCTCCGCGGCAAGTCTC 281
DB 181 AACAAAGCGCAACGCGACGATCCAAACGCGTGGTTCGGGAGCGCTCCGCGGCAAGTCTC 240

QY 282 GGCATCTCCACCGCGGCGTATCACCGCGTTCGACGCGCTCCGATCACTTCGCGCAC 341
DB 241 GGCATCTCCACCGCGGCGTATCACCGCGTTCGACGCGCTCCGATCACTTCGCGCAC 300

QY 342 GCGATGCGGCGAGCGCTTAACGGGCGATCATCCCGGTGACGCTCATCTCGGTGACCTGGCAA 401
DB 301 GCGATGCGGCGAGCGCTTAACGGGCGATCATCCCGGTGACGCTCATCTCGGTGACCTGGCAA 360

QY 402 ACCAAGTCGGGCGGACGCTACGAGGACGATGACATGCGGCGAGGACCCCGGCGGAA 461
DB 361 ACCAAGTCGGGCGGACGCTACGAGGACGATGACATGCGGCGAGGACCCCGGCGGAA 420

QY 462 TTCATGTGGATTCGGGGGCTTACCAACGAGATCAACTCCCGGAGGATGATACGCGGC 521
DB 421 TTCATGTGGATTCGGGGGCTTACCAACGAGATCAACTCCCGGAGGATGATACGCGGC 480

QY 522 CCGGGTTCGGCTCGTGGTGGCGCGGCTCAGATGCGGACAGCGTGGCGAGTACCTG 581
DB 481 CCGGGTTCGGCTCGTGGTGGCGCGGCTCAGATGCGGACAGCGTGGCGAGTACCTG 540

QY 582 TTTTCGGCGCGTTCGGGCTTCACTCGGTGCTGCGGCTGACGCTGGGCTGCTGGATA 641
DB 541 TTTTCGGCGCGTTCGGGCTTCACTCGGTGCTGCGGCTGACGCTGGGCTGCTGGATA 600

QY 642 GGTTCGTGCGGGTTCGTGATGGTGGCGCGGCTTCGCGTANGTGGCGTGGATGACGCTC 701
DB 601 GGTTCGTGCGGGTTCGTGATGGTGGCGCGGCTTCGCGTANGTGGCGTGGATGACGCTC 660

QY 702 ACCCGGGGCGAGCGGATGACGCGCGCCAGGTTCGGGTTCCTCGGCGGCGCTACGAG 761
DB 661 ACCCGGGGCGAGCGGATGACGCGCGCCAGGTTCGGGTTCCTCGGCGGCGCTACGAG 720

QY 762 ACGCGGTATGGGCTGACGCTGCGCGCGCGGCTGATCGCGAGAACCGTGTCACTGATG 821
DB 721 ACGCGGTATGGGCTGACGCTGCGCGCGCGGCTGATCGCGAGAACCGTGTCACTGATG 780

QY 822 ATTCTGATACGACCAACTCTTGGGGCAAAACACCGCGCGATCGGCTCAACGAGGCC 881
DB 781 ATTCTGATACGACCAACTCTTGGGGCAAAACACCGCGCGATCGGCTCAACGAGGCC 840

QY 882 GAATAAGCGGAGATGTGGGCGCCAAAGACGCGCGCGATGTTTGGCTACCGCGGCGACG 941
DB 841 GAATAAGCGGAGATGTGGGCGCCAAAGACGCGCGCGATGTTTGGCTACCGCGGCGACG 900

QY 942 GGCAGCGGACGCGGACGCTTGTCTCGGTTCCAGGAGGCGCGGAGATGACACGCGCGGT 1001
DB 901 GGCAGCGGACGCGGACGCTTGTCTCGGTTCCAGGAGGCGCGGAGATGACACGCGCGGT 960

QY 1002 GGGCTCTCTCGAGCAGCGCGCGCGGTTCGAGGAGGCTTCGACACCGCGCGCGGCAACCCAG 1061
DB 961 GGGCTCTCTCGAGCAGCGCGCGCGGTTCGAGGAGGCTTCGACACCGCGCGCGGCAACCCAG 1020

QY 1062 TTGATGAACAATGTGCGCGCGCGCGCTCAACAGCTGGGCGGCGCGGACGCGGACCCAG 1121
DB 1021 TTGATGAACAATGTGCGCGCGCGCGCTCAACAGCTGGGCGGCGCGGACGCGGACCCAG 1080

QY 1122 CTTTCTTCCAACTCGGCTGGCTGTGGAAGACGCTCTCGCGCATCGGCTCGCCCATCAGC 1181
DB 1081 CTTTCTTCCAACTCGGCTGGCTGTGGAAGACGCTCTCGCGCATCGGCTCGCCCATCAGC 1140

QY 1182 AACATGGTGTGATGCGGCGCAACCAACATGTCGATGACCAACTCGGCTGTCGATGACC 1241
DB 1141 AACATGGTGTGATGCGGCGCAACCAACATGTCGATGACCAACTCGGCTGTCGATGACC 1200

QY 1242 AACACCTTGAGCTCGATCTTGAAGGCTTTCCTCGGCGGCGCGCGGCGCGCTGCAA 1301
DB 1201 AACACCTTGAGCTCGATCTTGAAGGCTTTCCTCGGCGGCGCGCGGCGCGCTGCAA 1260

QY 1302 ACCCGGCGCGCAAAACGCGGTCCGGCGATGAGCTCGTGGGCGAGCTCGTGGGTTCCTCG 1361
DB 1261 ACCCGGCGCGCAAAACGCGGTCCGGCGATGAGCTCGTGGGCGAGCTCGTGGGTTCCTCG 1320

QY 1362 GGTCTGGCGGCTGGGCTGGCGCGCAACTTGGGTCGGGCGGCTCGCTCGGCTTCCTTCG 1421
DB 1321 GGTCTGGCGGCTGGGCTGGCGCGCAACTTGGGTCGGGCGGCTCGCTCGGCTTCCTTCG 1380

QY 1422 GTGCGCGAGGCTTGGGCGCGCGCGCAACCCAGGACAGTCAACCGCGGCGCGGCGCTCGCG 1481
DB 1381 GTGCGCGAGGCTTGGGCGCGCGCGCAACCCAGGACAGTCAACCGCGGCGCGGCGCTCGCG 1440

QY 1482 CTGACCGCTGACGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1541
DB 1441 CTGACCGCTGACGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1500

QY 1542 GTGGGCGAGATGGGCGCGCGGCGGCTGAGTGGGCTCAGTGGTGTGCTGCGGTTCGCGCG 1601
DB 1501 GTGGGCGAGATGGGCGCGCGGCGGCTGAGTGGGCTCAGTGGTGTGCTGCGGTTCGCGCG 1560

QY 1602 CGACCTTATGATGCGCGCATTTCTCCGCGAGCGCGGCGATATCGCGCGCGCGGCTTGTG 1661
DB 1561 CGACCTTATGATGCGCGCATTTCTCCGCGAGCGCGGCGATATCGCGCGCGCGGCTTGTG 1620

QY 1662 CAGGACCGGTTTCGCGGCTTCCCGCGCTCGCCCTCGACCGCTCGCGGATGTCGCGCAA 1721
DB 1621 CAGGACCGGTTTCGCGGCTTCCCGCGCTCGCCCTCGACCGCTCGCGGATGTCGCGCAA 1680

QY 1722 GTGGGCGGCGAGGTGTCATCAACATCAACCAAACTGGGCTTACAAACGCGCGTGGGCGCC 1781
DB 1681 GTGGGCGGCGAGGTGTCATCAACATCAACCAAACTGGGCTTACAAACGCGCGTGGGCGCC 1740

QY 1782 GGGACCGGCGATCGATCGATCCCAACGCTGTGCTGCTGACCAACCAACGATGATCGCG 1841
DB 1741 GGGACCGGCGATCGATCGATCCCAACGCTGTGCTGCTGACCAACCAACGATGATCGCG 1800

QY 1842 GCGCGCGCGGATCAATCAATGCGTTCAGCGGTTCGCGGCTTCGCGCGGCGGCGGCGGCG 1901
DB 1801 GCGCGCGCGGATCAATCAATGCGTTCAGCGGTTCGCGGCTTCGCGCGGCGGCGGCGGCG 1860

QY 1902 GTGGGCTATGACCGGCGGCGGATGTCGCGGTGCTGAGCTGCGCGGTGCGCGGTGCGCTG 1961
DB 1861 GTGGGCTATGACCGGCGGCGGATGTCGCGGTGCTGAGCTGCGCGGTGCGCGGTGCGCTG 1920

QY 1962 CCGTTCGCGCGGCGATCGGTCGCGGCTGAGCGGCTGAGCGGCTGCTGCGGCGGCGGCGGCG 2021
DB 1921 CCGTTCGCGCGGCGATCGGTCGCGGCTGAGCGGCTGAGCGGCTGCTGCGGCGGCGGCGGCG 1980
```

Qy 2022 ACCGTCGGCAGGCGGAACGCCCGTGGGTGCTTGGCAGGTTGGTCGGCTCGGCCAA 2081
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1981 ACCGTCGGCAGGCGGAACGCCCGTGGGTGCTTGGCAGGTTGGTCGGCTCGGCCAA 2040
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2082 ACCGTGCAGGCGTCGGATTTCGCTGACCGGTGCCGAAGAGACATTGAACGGTTGATCCAG 2141
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2041 ACCGTGCAGGCGTCGGATTTCGCTGACCGGTGCCGAAGAGACATTGAACGGTTGATCCAG 2100
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2142 TTCGATGCCCGCATCCAGCCCGGTGATTCCGGCGGGCGCGCTGTCACACGGCCTAGGACAG 2201
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2101 TTCGATGCCCGCATCCAGCCCGGTGATTCCGGCGGGCGCGCTGTCACACGGCCTAGGACAG 2160
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2202 GTGGTCGGTATGAACACGGCGCGCTCCTAG 2231
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 2161 GTGGTCGGTATGAACACGGCGCGCTCCTAG 2190
Db ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Search completed: July 4, 2004, 00:29:18
Job time : 1017 secs

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 3, 2004, 17:20:36 ; Search time 5931 Seconds

(without alignments)
11712.366 Million cell updates/sec

Title: US-09-597-796C-11

Perfect score: 2287
Sequence: 1 tctagaataattttgtttta.....ggntgtaaacgaacccgaaa 2287

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_esti:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pin:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_mam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_pbg:*

27: em_gss_vrl:*

28: gb_gss1:*

29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77.2	3.4	1628	29	CG757066 P052-2-A0
c 2	76.4	3.3	1406	29	CG756569 P051-4-B0
3	76.2	3.3	625	14	CD937289 OV.106112
c 4	76.2	3.3	935	29	CNS006XK AL066051 Drosophil

ALIGNMENTS

RESULT 1
CG757066
LOCUS P052-2-A02.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
DEFINITION genomic survey sequence.
ACCESSION CG757066 GI:37985257
VERSION CG757066.1
KEYWORDS GSS.
SOURCE Pristionchus pacificus
ORGANISM Pristionchus pacificus
Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
Neodiplogasteridae; Pristionchus.
REFERENCE 1 (bases 1 to 1628)
AUTHORS Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K., Buntjer,J., van der Meulen,M. and Sommer,R.J.
TITLE An integrated physical and genetic map of the nematode Pristionchus pacificus
JOURNAL Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE 22835951
PUBMED 12884007
COMMENT Contact: Sommer RJ
Evolutionary Biology
Max-Planck-Institute for Developmental Biology
Spemannstr. 37-39, Tuebingen D-72076, Germany
Tel: 00497071601371
Fax: 00497071601498
Email: ralf.sommer@tuebingen.mpg.de

5	75	3.3	611	13	BU102503
6	75	3.3	611	13	CA064610
7	74.6	3.3	925	29	CNS0081P
c 8	73.6	3.2	1516	12	EGS0984
9	72.8	3.2	935	29	CNS006XK
10	72.6	3.2	949	29	AG171092
c 11	72.4	3.2	925	29	CNS0091P
c 12	72	3.1	1798	29	AG171124
c 13	71.8	3.1	1956	29	CG754548
c 14	71.2	3.1	880	29	CG458121
c 15	70.8	3.1	1160	12	EG844853
c 16	70	3.1	550	14	CB869131
c 17	70	3.1	897	29	AG046101
c 18	69.6	3.0	1061	13	EX391246
19	69.4	3.0	932	29	CNS0072Q
20	69.4	3.0	1116	28	BZ569478
21	68.8	3.0	1203	29	CNS015Y4
c 22	68.6	3.0	1079	29	AG080398
23	68.4	3.0	862	29	AG036151
c 24	68.2	3.0	1218	13	BQ876717
c 25	68	3.0	1151	14	CK210749
c 26	67.8	3.0	289	29	U82114
27	67.8	3.0	1057	13	EX349688
c 28	67.6	3.0	750	29	AG149322
c 29	67.6	3.0	1000	13	EX407619
30	67.4	2.9	982	13	EX415111
c 31	67.2	2.9	807	14	CF149720
32	67.2	2.9	821	28	AG063084
33	67.2	2.9	1474	28	BZ569821
34	67	2.9	641	13	CA109290
c 35	66.8	2.9	932	29	CNS0072Q
c 36	66.6	2.9	924	13	BU187694
c 37	66.6	2.9	1121	29	AG062320
c 38	66.6	2.9	1798	29	AG171124
c 39	66.2	2.9	868	29	AG061755
c 40	66	2.9	742	29	CC688263
c 41	66	2.9	851	14	CB908312
c 42	66	2.9	978	29	CC735401
c 43	66	2.9	1075	29	AG073983
c 44	66	2.9	1200	13	EX456467
c 45	66	2.9	1288	13	BQ678719

CG757066 1628 bp DNA linear GSS 24-OCT-2003
P052-2-A02.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
genomic survey sequence.

CG757066 GI:37985257

CG757066.1

GSS.

Pristionchus pacificus

Pristionchus pacificus

Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;

Neodiplogasteridae; Pristionchus.

1 (bases 1 to 1628)

Srinivasan,J., Sinz,W., Jesse,T., Wiggers-Perebolte,L., Jansen,K.,

Buntjer,J., van der Meulen,M. and Sommer,R.J.

An integrated physical and genetic map of the nematode Pristionchus

pacificus

Mol. Genet. Genomics 269 (5), 715-722 (2003)

22835951

12884007

Contact: Sommer RJ

Evolutionary Biology

Max-Planck-Institute for Developmental Biology

Spemannstr. 37-39, Tuebingen D-72076, Germany

Tel: 00497071601371

Fax: 00497071601498

Email: ralf.sommer@tuebingen.mpg.de

Db 83 CAGGAGCGCGCGCGTGCAGGTGGTCTCGCGGAGCGCGGGAAGCGGTGCGCGGTG 142
 QY 558 TGGGACACGCGTGGCGAGTGACCTGTTTTCGCGCGCGTGGCGGTTCAGTCCGTTGCTCGG 617
 Db 143 GTGGCTTGGAGGTGGTTCAGCGCGGAGGAGGAGTCCGTTGGTGGTGGTGGCGAGG---AG 199
 QY 618 GGTCTGACGCGTGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 677
 Db 200 GTTCAGGTGGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCGTGGCG 259
 QY 678 CCGTATGTTGGCGTGGATGAGCGTCAACCGCGCGGCGAGCGCGAGTGCAGCCGCGCGCGGTC 737
 Db 260 GAGGAGGTGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 319
 QY 738 CGGTTGCTGGCGCGCGCTACGAGACGCGCGTATGGGCTGACGCGTGGCGCGCGCGCGTGCATC 797
 Db 320 GCCTTGGAGGAGGTGTCGGATCAGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 379
 QY 798 GCCGAGAACCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 857
 Db 380 GCGTGGCTTGGAGGAGGTGGCGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 439
 QY 858 CCGCGCATCGCGTCAACGAGCGCGGCGAATACGCGGAGATGTGGCGCGCGCGCGCGCGCGG 917
 Db 440 GCTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 499
 QY 918 ATGTTTGGCTACGCGCGCGGCGAGCGCGACCGCGCGAGCGCGTGGTGGTGGTGGTGGTGGTGG 977
 Db 500 GTGAGGAGCTCANGCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 559
 QY 978 GCGCGCGGAGATGACGAGCGCGGTGGGCTC 1007
 Db 560 GAGCGCGGAGGCGGAGCG 589

RESULT 4
 CNS006XK/c 935 bp DNA linear GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence T7 end of BAC #
 DEFINITION BACR14N09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 ACCESSION AL066051
 VERSION AL066051.1 GI:4945019
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 935)
 AUTHORS Direct Submission
 TITLE Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazuhiro Osogawa and Aaron Mammox in Peter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's pi and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
 FEATURES Location/Qualifiers
 1. .935

ORIGIN
 Query Match 3.3%; Score 76.2; DB 29; Length 935;
 Best Local Similarity 28.8%; Pred. No. 0.29;
 Matches 122; Conservative 114; Mismatches 187; Indels 0; Gaps 0;
 QY 1421 GGTGCGCAGCGTGGCG 1480
 Db 927 GSGSSSSGSGCGSG 868
 QY 1481 GGTGACCAAGCTGACAGCG 1540
 Db 867 GCGSCGSG 808
 QY 1541 GGTGCGCGCAGATGGCG 1600
 Db 807 GSGCG 748
 QY 1601 GCGACCCCTATGTGATCGCGCATTCCTCGCGCAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1660
 Db 747 GSGCGSG 688
 QY 1661 GCGAGACCGGTTCG 1720
 Db 687 GSGCG 628
 QY 1721 AGTGGGCGCAGGTGTCAACATCAACACCAACTGGGCTACACCAACCGCGCGCGCGCGCG 1780
 Db 627 GCGSCG 569
 QY 1781 GCGGACCGCGCATCGTATCGATCCCAACGGTGTGCTGTGCTGACCAACCAACCGCGCGCGCG 1840
 Db 567 GCGMCGRAGSGKMSAGSG 508
 QY 1841 GGG 1843
 Db 507 GBK 505
 RESULT 5
 BUI02503
 LOCUS SCCAD1001A12.g Saccharum officinarum mRNA (Nogueira,F.T.S)
 DEFINITION Saccharum officinarum cDNA, mRNA sequence.
 ACCESSION BUI02503
 VERSION BUI02503.1 GI:32813826
 KEYWORDS EST.
 SOURCE Saccharum officinarum
 ORGANISM Saccharum officinarum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Saccharum.
 REFERENCE 1 (bases 1 to 611)
 AUTHORS Nogueira,F.T.S., de Rosa,V.E. Jr., Menossi,M., Ulian,E.C. and Arruda,P.
 TITLE RNA expression profiles and data mining of sugarcane response to low temperature
 JOURNAL Plant Physiol. 132 (4), 1811-1824 (2003)
 MEDLINE 22795309
 PUBMED 12913139
 COMMENT Contact: Nogueira FTS
 Bioinformatics Lab
 Organization for Nucleotide Sequencing and Analysis
 C.P. 6176, Campinas, SP 13083-970, Brazil
 Tel: 55 19 37881101
 Fax: 55 19 37881089
 Email: tebaldi@unicamp.br.


```

FEATURES
  source
    Location/Qualifiers
      1. .611
      /organism="Saccharum officinarum"
      /mol_type="mRNA"
      /db_xref="taxon:4547"
      /clone_lib="Saccharum officinarum mRNA (Nogueira, F.T.S)"
      /db_xref="taxon:4547"
      /clone_lib="Saccharum officinarum mRNA (Nogueira, F.T.S)"

ORIGIN
  Query Match      3.3%; Score 75; DB 13; Length 611;
  Best Local Similarity 47.8%; Pred. No. 0.39;
  Matches 253; Conservative 0; Mismatches 270; Indels 6; Gaps 1;

  QY 566 CTTGGGAGTACCTGTTTTCGGCGCGCTCGCGGTTCAGTCGGTGTCTGCGGTCTGAC 625
  Db 9 CATGGCGCGCTTGGAGGTGCGCTTGGAGGTGGCGGCGGCGGTGCTCTCGTGGTGGT 68
  QY 626 GGTGGGTCTGTGATAGTTTCGTGGCGGTCTGAT-----GGTGGCGCGGCTCGCC 679
  Db 69 GCAGGCGAGAGGCGCGCTTCGCGGTGGAGGTGGTGCAGGCGGAGCGCGCGCTTGGT 128
  QY 680 GTATGTCGGCTGATGAGCGCTCACCGCGGCGAGCGCGGCTGACGCGCGCGCGGTCG 739
  Db 129 GGTGGTCCCGCGGTGCTTTGGAGGTGGTGCAGGCGGAGGAGGCGCTTGTGGCGGT 188
  QY 740 GTTTGTCTGCGCGCGCTTACGAGAGCGGCTATGGGCTGACGTCGCCCGCGCGGTGATCG 799
  Db 189 GCGGCTGAGGAGGTGGACTAGGCGGTGGTGGCGGCGCTTGGCGGGGCGCAGGA 248
  QY 800 CGAAGACCGTGTCACTGATGATTCGATAGCGACCACTCTTGGGGCAAAACACCCC 859
  Db 249 GGAGTACTCGCGGTGGCTTGGTGGTGGTGGCGCGGTTCGAGGAGGCGCAAGCGCG 308
  QY 860 GCGCATCGCGTCAACGAGGCGGAATACGCGCGAGATGTGGGCGGCAAGACGCGCGCAT 919
  Db 309 GGAGGAGAGGCGCTTGGTGGCGGTGGCGGTGGAGGAGGTGACTAGCGGTGGTGGCGGT 368
  QY 920 GTTTGGCTACGCGCGCGCGCGAGCGCGAGCGCGAGCGGTGCTGCGGTTCGAGGAGCG 979
  Db 369 GCGCGCGCGGAGGTGGTGGCGGAGCGGCGGCTTTTGGCGAGGCAAGAGGCGCGC 428
  QY 980 GCGCGAGATGACACGCGCGGTGGCTCTCGAGAGCGCGCGCGCTCGAGGAGGCGCTC 1039
  Db 429 TTCGAGGTGGCGCGCGCGAGCGGTGGTGGCGGTGGCGGTTCGAGGAGGCAAGGCGCG 488
  QY 1040 CGACACCGCGCGCGCGAGCAACGATGATGAACAATGTGCCCCAGCGCTG 1088
  Db 489 GGTTTTGGCGCGCGCAAGGAGGCGGTTCGCGGAGGCGCGAGGTGGTG 537

  RESULT 7
  CNS0091P
  LOCUS
  DEFINITION
    Drosophila melanogaster genome survey sequence TET3 end of BAC #
    BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit
    fly), genomic survey sequence.

  925 bp DNA linear GSS 03-JUN-1999
  
```

```

FEATURES
  source
    Location/Qualifiers
      1. .611
      /organism="Saccharum officinarum"
      /mol_type="mRNA"
      /db_xref="taxon:4547"
      /clone_lib="Saccharum officinarum mRNA (Nogueira, F.T.S)"
      /db_xref="taxon:4547"
      /clone_lib="Saccharum officinarum mRNA (Nogueira, F.T.S)"

ORIGIN
  Query Match      3.3%; Score 75; DB 13; Length 611;
  Best Local Similarity 47.8%; Pred. No. 0.39;
  Matches 253; Conservative 0; Mismatches 270; Indels 6; Gaps 1;

  QY 566 CTTGGGAGTACCTGTTTTCGGCGCGCTCGCGGTTCAGTCGGTGTCTGCGGTCTGAC 625
  Db 9 CATGGCGCGCTTGGAGGTGCGCTTGGAGGTGGCGGCGGCGGTGCTCTCGTGGTGGT 68
  QY 626 GGTGGGTCTGTGATAGTTTCGTGGCGGTCTGAT-----GGTGGCGCGGCTCGCC 679
  Db 69 GCAGGCGAGGAGGCGCGCTTCGCGGTGGAGGTGGTGCAGGCGGAGCGCGCGCTTGGT 128
  QY 680 GTATGTCGGCTGATGAGCGCTCACCGCGGCGAGCGCGAGTGCAGCGCGCGCGCGTCCG 739
  Db 129 GGTGGTCCCGCGGTGCTTTGGAGGTGGTGCAGGCGGAGGAGGCGCTTGTGGCGGT 188
  QY 740 GTTTGTCTGCGCGCGCTTACGAGAGCGGTATGGGCTGACGCGCGCGCGGTGATCGC 799
  Db 189 GCGGCTGAGGAGGTGGACTAGGCGGTGGTGGCGGTGGTGGCGGCGGCGCGCGCAT 248
  QY 800 CGAAGACCGTGTCACTGATGATTCGATAGCGACCACTCTTGGGGCAAAACACCCC 859
  Db 249 GGAGTACTCGCGGTGGCTTGGTGGTGGTGGCGGCGGTTCGAGGAGGCGCAAGCGCG 308
  QY 860 GCGCATCGCGTCAACGAGGCGGAATACGCGCGAGTGTGGGCGGCAAGCGCGCGGTGATCG 919
  Db 309 GCGGCTGAGGAGGTGGTGGCGGAGCGCGGCTTTGGCGGAGGCAAGAGGCGCGC 428
  QY 920 GTTTGTCTGCGCGCGCTTACGAGAGCGGCTATGGGCTGACGTCGCCCGCGCGGTGATCG 979
  Db 369 GCGGCTGAGGAGGTGGTGGCGGAGCGCGGCTTTGGCGGAGGCAAGAGGCGCGC 428
  QY 980 GCGGAGATGACACGCGCGGTGGCTCTCGAGAGCGCGCGCGGTTCGAGGAGGCGCTC 1039
  Db 429 TTCGAGGTGGCGCGCGCGAGCGGTGGTGGCGGTGGCGGTTCGAGGAGGCGCAAGCGCG 488
  QY 1040 CGACACCGCGCGCGCGAGCAACGATGATGAACAATGTGCCCCAGCGCTG 1088
  Db 489 GGTTTTGGCGCGCGCAAGGAGGCGGTTCGCGGAGGCGCGAGGTGGTG 537

  CA064610 611 bp mRNA linear EST 23-SEP-2003
  LOCUS
  DEFINITION
    SCCAD1001A12.g Adl Saccharum officinarum cDNA clone SCCAD1001A12
    5', mRNA sequence.
  ACCESSION
    CA064610
  VERSION
    CA064610.1 GI:34916134
  KEYWORDS
    EST.
  SOURCE
    Saccharum officinarum
  ORGANISM
    Saccharum officinarum
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
    Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
    clade; Panicoideae; Andropogoneae; Saccharum.
    1 (bases 1 to 611)
    Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
    The libraries that made SUCEST
    Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
    Contact: Arruda P
    Centro de Biologia Molecular e Engenharia Genetica
    Universidade Estadual de Campinas
    Caixa Postal 6010, 13083-970, Campinas SP, Brazil
    Tel: 55 19 3788 1137
    Fax: 55 19 3788 1089
    Email: parruda@unicamp.br
  
```



```
QY 1760 CTACAAACACCGGCGGCGGACCGGAT--CGTCATCATGCCAACGGTGTGCTG 1817
Db 785 CNGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 726
QY 1818 CTGACCAACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1877
Db 725 CCCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 666
QY 1878 GGCAGCACTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1935
Db 665 GCGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 606
QY 1936 TGCAGTGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1995
Db 605 CCCCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 546
QY 1996 GTGAGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2055
Db 545 CCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 487
QY 2056 CTGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2115
Db 486 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 427
QY 2116 AAGAGACATTTGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2175
Db 426 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 367
QY 2176 GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2208
Db 366 CCCCAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 334
```

```
RESULT 9
CNS006XK
LOCUS
DEFINITION
Drosophila melanogaster genome survey sequence 17 end of BAC #
BAC14N09 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION
AL066051
VERSION
AL066051.1 GI:4945019
KEYWORDS
GSS.
SOURCE
Drosophila melanogaster (fruit fly)
ORGANISM
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE
1 (bases 1 to 935)
Genoscope.
AUTHORS
Direct Submission
TITLE
Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
COMMENT
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosier in Pictet de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila\_bac.htm.
FEATURES
source
1..935
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
```

```
/clone="BAC14N09"
/clone_lib="RPCI-98"
/notes="end : T7"

ORIGIN
Query Match 3.2%; Score 72.8; DB 29; Length 935;
Best Local Similarity 35.6%; Pred. No. 0.88;
Matches 116; Conservative 78; Mismatches 130; Indels 2; Gaps 1;

QY 1275 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1334
Db 611 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 670
QY 1335 TCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1394
Db 671 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 728
QY 1395 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1454
Db 729 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 788
QY 1455 GTACCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1514
Db 789 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 848
QY 1515 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1574
Db 849 SGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 908
QY 1575 CTCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1600
Db 909 CCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 934
```

```
RESULT 10
AG171092
LOCUS
DEFINITION
Pan troglodytes DNA, clone: RP43-040E09.TJ, genomic survey
sequence.
ACCESSION
AG171092
VERSION
AG171092.1 GI:16700770
KEYWORDS
GSS.
SOURCE
Pan troglodytes (chimpanzee)
ORGANISM
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE
1
AUTHORS
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE
BAC end sequences of Library RPCI-43
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 949)
AUTHORS
Fujiyama,A., Hattori,M., Toyoda,A., Taylor,T.D., Yada,T.,
Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE
Direct Submission
JOURNAL
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
1-7-22 Suehiro-chou,Tsukumi-Ku, Yokohama, Kanagawa 230-0045, Japan
(E-mail:chimbases@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
Tel.:81-45-503-9111, Fax:81-45-503-9170)
COMMENT
Clones are derived from the chimpanzee BAC library RPCI-43 This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI
FEATURES
source
1..949
Location/Qualifiers
/organism="Pan troglodytes"
/mol_type="genomic DNA"
```


sequence.
ACCESSION AG171124
VERSION AG171124.1 GI:16700802
KEYWORDS GSS.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
1
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
BAC end sequences of Library RPCI-43
Unpublished
2 (bases 1 to 1798)
Fujiyama, A., Hattori, M., Toyoda, A., Taylor, T.D., Yada, T.,
Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
Submitted (02-AUG-2001) Asao Fujiyama, The Institute of Physical
and Chemical Research (RIKEN), Genomic Sciences Center (GSC); Japan
1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0445, Japan
(E-mail: chimpansegsc.riken.go.jp, URL: http://hgp.gsc.riken.go.jp/,
Tel: 81-45-503-9111, Fax: 81-45-503-9170)
Clones are derived from the chimpanzee BAC library RPCI-43. This BAC
end was generated during the R&D process and may have higher chance
of clone tracking errors.
PRIMERS
Sequencing: TJ
LIBRARY
Vector : pBACe3.6
R.Site 1 : EcoRI
R.Site 2 : EcoRI.
FEATURES
source
1..1798
Location/Qualifiers
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/clone="RP43-040F09.TJ"
/sex="male"
/cell_type="lymphocytes"
/clone_lib="RPCI-43 Chimpanzee Male BAC Library"
ORIGIN
Query Match 3.1%; Score 72; DB 29; Length 1798;
Best Local Similarity 37.6%; Pred. No. 1.3;
Matches 554; Conservative 0; Mismatches 910; Indels 9; Gaps 5;
QY 612 GTCTGGGTCTACGGTGGGGTCTGGATAGTTTCTGTGGGGGTCTGATGGTGGCGG 671
DB 1737 GCCCGGGGNNCGCGCGCNCNNCGCGGCGCGGCGCGGCGCGGCGCGGCGCGG 1678
QY 672 GCCTCGCCGTATGTGCGGTGATGAGCTACCGGGGAGCGCGAGTGAACCCGCC 731
DB 1677 CG 1618
QY 732 CAGGTCCGGGTGCTGCGCGCGCCTACGAGACGGGTATGGGCTGACGGTCCCGCGG 791
DB 1617 CGNNGCGCGCGGGGCGCGCGGCGCGGCGCGGCGCGCGCGCGCGCGCGCGCG 1559
QY 792 GTGATCCCGAGAACCGTCTGAATGATGATCTGATAGACCAACCTCTTGGGGCAA 851
DB 1558 CCGGNGNNGNCGGNGNCGGNGCGGNGCGGCGCGGCGCGGCGGCGGCGGCGG 1499
QY 852 AACACCCCGCGATCGCGGTCAACGAGCGCGAATACGGCGAGATGTGGGCCCAAGACGC 911
DB 1498 NNGNGNGCG 1439
QY 912 GCGCGATGTTGGTACGCGCGGAGCGCGGAGCGGAGCGGAGCGGAGCGGAGCGG 971
DB 1438 GNCNCG 1379
QY 972 GAGGAGCGCGGAGATGACGAGCGGCGGTGGGCTCTCGAGCAGCGCGCGCGTCCAG 1031
DB 1378 GCG 1319

```

RESULT 13
CG754548      1956 bp      DNA      linear      GSS 24-OCT-2003
LOCUS        P050-1-A11.za Ppa EcoRI BAC Library Pristionchus pacificus genomic,
DEFINITION   genomic survey sequence.
ACCESSION    CG754548
VERSION      CG754548.1 GI:37980151
KEYWORDS     GSS.
SOURCE       Pristionchus pacificus
ORGANISM     Pristionchus pacificus
              Eukaryota; Metazoa; Nematoda; Chromadorea; Diplogasterida;
              Neodiplogasteridae; Pristionchus.
REFERENCE    1 (bases 1 to 1956)
AUTHORS      Srinivasan, J., Sinz, W., Jesse, T., Wiggers-Perebolte, L., Jansen, K.,
              Buntjer, J., van der Meulen, M. and Sommer, R.J.
TITLE        An integrated physical and genetic map of the nematode Pristionchus
              pacificus
JOURNAL      Mol. Genet. Genomics 269 (5), 715-722 (2003)
MEDLINE      22835951
PUBMED       12884007
COMMENT      Contact: Sommer RJ
              Evolutionary Biology
              Max-Planck-Institute for Developmental Biology
              Spemannstr. 37-39, Tuebingen D-72076, Germany
              Tel: 00497071601371
              Fax: 00497071601498
              Email: ralf.sommer@uebingen.mpg.de
              Class: BAC ends
              Location/Qualifiers
              1..1956
              /organism="Pristionchus pacificus"
              /mol_type="genomic DNA"
              /strain="California"
              /db_xref="taxon:54126"
              /clone_lib="Ppa EcoRI BAC Library"
              /note="The library was generated by a partial digest of
              the genomic DNA with EcoRI and cloning into the BAC
              vector."

FEATURES             source
Query Match          3.1%; Score 71.8; DB 29; Length 1956;
Best Local Similarity 43.3%; Pred. No. 1.4;
Matches 457; Conservative 0; Mismatches 583; Indels 16; Gaps 5;

Qy 516 GCGGCGCCGGTTCGGCTCTGCTGGTGGCGCGGCTCAGATGTGGACACGCTGGCGAGT 575
Db 880 GNNNGNNNGGGNNNNNGGGGGGGGGGGGNGGNGGNNNGGGGGCCCGCCCGCGGNCN 939

Qy 576 GACCTGTTTCGGCGCGTTCAGTCTGGTCTGCTGGGCTGACGCTGGGGTTCG 635
Db 940 GNNNGNNNGGGGGGGGGGGGNCGCCCGCGGGGGGNGCCGGCGCCGCGCGCGCC 999

Qy 636 TGGATAGTTTCGTCGGCGGGTCTGATGGTGGCGCGGCTCGCCGATGTGGCGTGGATG 695
Db 1000 GGCGCNCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG--- 1056

Qy 696 AGCTACCGCGGGGAGCGAGCTGACCGCGCGCCAGCTCGCGGTTGTCGGCGCGCC 755
Db 1057 ---GCCCGCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1113

Qy 756 TACGAGACGGCGTATGGGCTGAGGCTGCCCGCGCGCGCGCGCGCGCGCGCGCGCG 815
Db 1114 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1173

Qy 816 CTGATGATTTGATAGGACCAACCTCTTGGGGGCAAAACACCCCGGAGTCGGGTCAAC 875
Db 1174 CGGGCCCGCGCGGGGGG---CGGCCGNGCGGGGGCGGCGGCGCGCGCGCGCGCGCG 1231

Qy 876 GAGCCGGAATACGCGGAGATGTGGGCGCAAGACGCGCGCGCGCGATGTTGGCTACGCGCG 935
Db 1232 GCGNCGGGCGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGG 1291

```

```

Qy 936 GCGACGGCGACGGCGACCGCGAC-----GTTGCTGCGGTTTCGAGAGGCGCGCGAGATGA 990
Db 1292 GCNCGNGGCGNCCCGGGCGCGCCCGCGGGCGGGGCGCGCGGGCGGGGCGGGGCGG 1351

Qy 991 CCAGCGCGGTGGGTCTCTCAGAGCAGCGCGCGCGGTCTGAGAGGCTCTCCACACCGCG 1050
Db 1352 CCGGGGGGGGGGGCGCGCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1411

Qy 1051 CGGCGAACCACTTGATGAACAATGTCGCCAGGCGCTGCAACAGCTGCGCCCGACCCACGC 1110
Db 1412 GCCGCGCGCGNGGGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1471

Qy 1111 AGGCGACCAACCCCTTCTTCCAAAGTGGGTGGCTGTGGAAGACGGTCTCGCGGATCGGT 1170
Db 1472 GCGCGCGCGGGCGCGCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1531

Qy 1171 CGCGATCAGCAACATGCTGTCGATGGCCAAACACACATGTCGATGACCAACTCGGGTG 1230
Db 1532 CCGGGGGCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1591

Qy 1231 TGTGATGACCAACACCTTTGAGCTCGATGTTTGAAGGCTTTTGTCCGCGCGCGCGCCGCC 1290
Db 1592 CCGGGCGCGCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1651

Qy 1291 AGGCGGTGCAAAACCGCGCGCGCAAAACCGGGTTCGGGGCGATGAGCTCGTGGGAGCTCGC 1350
Db 1652 GGGCGGGCGCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1709

Qy 1351 TGGGTTCTTTCGGGTCTGGGGCGTGGGTGGCGCGCAACTTGGGTGCGGCGCGCGCTCGGTCG 1410
Db 1710 CGGGCGGGGGCGCGCGCGCGCGCGCGGGGGGGCGCGCGGGGGGGGGGGGGGGGG 1768

Qy 1411 GTTCGTTGTCGGTGGCGAGGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1470
Db 1769 GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1828

Qy 1471 GGGCGCTGCGCTGACGAGCTGACCGCGCGCGGAAAGAGGCGCGCGCGCGCGCGCGCGCGCG 1530
Db 1829 GGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1888

Qy 1531 CGGGCTGCGCGTGGGGGAGATGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1566
Db 1889 GCGCGCGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1924

RESULT 14
CG458121/c      880 bp      DNA      linear      GSS 17-SEP-2003
LOCUS        PUFXV40TBC ZM 0.6_1.0.KB Zea mays genomic clone ZMMB7a075H08,
DEFINITION   genomic survey sequence.
ACCESSION    CG458121
VERSION      CG458121.1 GI:34843121
KEYWORDS     GSS.
SOURCE       Zea mays
ORGANISM     Zea mays
              Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
              Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
              clade; Panicoideae; Andropogoneae; Zea.
              1 (bases 1 to 880)
REFERENCE    Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
              Resnick, A., Fraser, C.M., Yuan, Y., San Miguel, P., Ma, J. and
              Bennetzen, J.
              Maize Genomics Consortium
              Unpublished (2003)
JOURNAL      Other GSSs: PUFXV40TDC
COMMENT      Contact: Cathy Whitelaw
              TIGR
              9712 Medical Center Drive, Rockville, MD 20850, USA
              Tel: 301-838-5843
              Fax: 301-838-0208
              Email: whitelaw@tigr.org
              Seq primer: TR
              Class: Sheared ends.

```

FEATURES	Location/Qualifiers	
	source	
	1..880	
	/organism="Zea mays"	
	/mol_type="genomic DNA"	
	/strain="B73"	
	/db_xref="taxon:4577"	
	/clone="ZMMBta0757H08"	
	/clone_lib="ZM 0.6-1.0 KB"	
	/note="Vector: pCR4-TOPO; Site 1: EcoRI; Site 2: EcoRI; 0.6-1.0 kb high Cot selected genomic DNA library"	
ORIGIN	Query Match 3.1%; Score 71.2; DB 29; Length 880;	
	Best Local Similarity 45.9%; Pred. No. 1.5; Mismatches 288; Indels 0; Gaps 0;	
	Matches 244; Conservative 0; Mismatches 288; Indels 0; Gaps 0;	
	QY	536 GCTGTGCGCGCGCTCAGATGTGGACAGGTGGCGAGTACCTGTTTGGCGCGCTC 595
	DB	682 GTTGTGCGGGGAGGATATGGCGAGCGGTGGCGGTGGCGGTGGCGGTGGT 623
	QY	596 GCGGTTTCAGTGGTGTGGGTCTGACGTGGGTGGGTGGGTGGGTGGGTGG 655
	DB	622 GGTGGCTTGGGACAGGTGTAGGGTTGGAATTGGATTGGAGAGGAATGTGGGT 563
	QY	656 TCTGATGGTGGCGGCGCTCGCGGTATGTGGGTGGATAGCGTACCGCGGGGAGCG 715
	DB	562 GCAGTGGTGGAGCGCGCTCGGTGGTGGAGCTGGAGTGGTGGTGGTGGTGG 503
	QY	716 CGAGCTGACCGCGCCAGGTGGGTCTGCGCGCGCTACGAGCGCGTATGGGCT 775
DB	502 CTTGTGCGGTGGAGGAGGTGTTATGGCGAGGTCTGGCGAGGCGCGCGCTTGG 443	
QY	776 GACGTGCCCCCGCGGTGATCGCGAGAACCGTGTGAACCTGATGATTTCTGATAGCGAC 835	
DB	442 GGTGTTTCGGTGGAGGTCTCGCGGTGTGAAGCGGAGGAGCGCGGTCTGGTGTGT 383	
QY	836 CAACCTTTGGGGCAAAACACCCCGCGATCGCGTCAACGAGGCGGATATCGCGGAGAT 895	
DB	382 GCGCGTGGAGCGGTGGTCTTGTGGCGGAGGAGGCTGGTGGGGAGCAGGTGGCGGT 323	
QY	896 GTGGCGCCAAAGACCGCGCGGATGTTTGGCTACGCGCGCGACGCGCGACGCGCGCG 955	
DB	322 GCGCGCTCTGGCGAGGCGCGGTGACAGAGTGGCGCTGGAGCGGCTATGGCGGAGGT 263	
QY	956 GAGCTTGTGCGTTTCGAGGAGGCGCGGAGATGACACGCGGGGTGGGTCTCTCGAGCA 1015	
DB	262 GCGGTGTCAGAGGCGCGGTATGTTGGAGAGCTGGTGGTGGATACGCTGGAGGTGACGGA 203	
QY	1016 GGCGCGCGGTGTCAGAGGCTCCGACACCGCGCGCGGCGAACAGTTGATG 1067	
DB	202 GGTGGCGCGGAGCGGCTACGCGGAGGAGCTGGTGGTGGTGGTGGTGGTGGTGGTGG 151	
RESULT 15		
BG844853/c 1160 bp mRNA linear EST 29-MAY-2001		
LOCUS 1024008A08.y2 C. reinhardtii CC-1690, normalized, Lambda Zap II		
DEFINITION Chlamydomonas reinhardtii cDNA, mRNA sequence.		
ACCESSION BG844853		
VERSION BG844853.1 GI:14226037		
KEYWORDS EST.		
SOURCE Chlamydomonas reinhardtii		
ORGANISM Chlamydomonas reinhardtii		
REFERENCE 1 (bases 1 to 1160)		
AUTHORS Grossman,A., Davies,J., Federspiel,N., Harris,E., Lefebvre,P., McDermott,J.P., Silflow,C., Stern,D. and Surzycki,R.		
TITLE Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 2		
JOURNAL Unpublished (2000)		
COMMENT Contact: Charles Hauser		

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:39:00 ; Search time 27.0437 Seconds
(without alignments)
2747.774 Million cell updates/sec

Title: US-09-597-796c-26

Perfect score: 1306

Sequence: 1 VAWMSVTAGQALTAQVRV.....YHRDGGKYXSGRRNGGPA 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_29Jan04.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1304	99.8	263	2 AAW32379	Mycobacte
2	1304	99.8	263	2 AAW32447	Mycobacte
3	1304	99.8	263	2 AAW64317	Mycobacte
4	1304	99.8	263	2 AAW61680	M. tuberc
5	1304	99.8	263	2 AAY32062	Mycobacte
6	1304	99.8	263	2 AAY38984	M. tuberc
7	1304	99.8	263	2 AAY39121	M. tuberc
8	1304	99.8	263	5 AAE29706	Mycobacte
9	1304	99.8	263	5 AAE17570	Mycobacte
10	1304	99.8	358	5 AAU74591	Antigenic
11	1187	90.9	391	2 AAW32381	Mycobacte
12	1187	90.9	391	2 AAW32449	Mycobacte
13	1187	90.9	391	2 AAW64335	Mycobacte
14	1187	90.9	391	2 AAW81702	M. tuberc
15	1187	90.9	391	2 AAY04778	Mycobacte
16	1187	90.9	391	2 AAY38989	M. tuberc
17	1187	90.9	391	2 AAY39132	M. tuberc
18	1187	90.9	391	4 AAU01888	M. tuberc
19	1187	90.9	391	5 AAE29707	Mycobacte
20	1187	90.9	391	5 AAE17571	Mycobacte
21	1187	90.9	596	2 AAY32070	Mycobacte
22	1187	90.9	596	5 AAE29710	Mycobacte
23	1187	90.9	596	5 AAE17574	Mycobacte
24	1187	90.9	599	5 AAU74599	Antigenic
25	1187	90.9	600	2 AAY32068	Mycobacte

26	1187	90.9	600	5 AAU74597	Antigenic
27	1187	90.9	723	7 ADA26354	Mycobacte
28	1187	90.9	723	4 AAO22142	Ral2-H9-3
29	1187	90.9	723	5 AAE29703	Mycobacte
30	1187	90.9	723	5 AAE17573	Mycobacte
31	1187	90.9	729	7 ADA26374	Mycobacte
32	1187	90.9	744	4 AAU01902	M. tuberc
33	1187	90.9	788	4 AAU01903	M. tuberc
34	1187	90.9	813	7 ADA26367	Mycobacte
35	1187	90.9	815	4 AAU01904	M. tuberc
36	1187	90.9	825	7 ADA26365	Mycobacte
37	1187	90.9	875	7 ADA26365	Mycobacte
38	1187	90.9	930	5 AAE29731	Mycobacte
39	1187	90.9	930	7 ADA26364	Mycobacte
40	1187	90.9	1010	7 ADA26356	Mycobacte
41	1187	90.9	1016	7 ADA26370	M. bovis
42	1187	90.9	1022	7 ADA26369	Mycobacte
43	1187	90.9	1154	7 ADA26368	Mycobacte
44	1182	90.5	394	2 AAY04779	Mycobacte
45	1182	90.5	729	5 AAE29708	Mycobacte

ALIGNMENTS

RESULT 1

AAW32379
ID AAW32379 standard; protein; 263 AA.
XX
AC AAW32379;
XX
DT 13-JAN-1998 (first entry)
XX
DE Mycobacterium tuberculosis antigen TbH-9.
XX
KW Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
XX skin testing; M.tuberculosis.
XX
OS Mycobacterium tuberculosis.
XX
FH Key Location/Qualifiers
FT Misc-difference 254
FT /note= "Any amino acid"
XX
FN WO9709429-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US014675.
XX
PR 01-SEP-1995; 95US-00523435.
PR 22-SEP-1995; 95US-00332136.
PR 22-MAR-1996; 96US-00620280.
PR 05-JUN-1996; 96US-00658800.
PR 12-JUL-1996; 96US-00680573.
XX
PA (CORI-) CORIXA CORP.
XX
PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedwick TH, Twardzik DR;
XX
DR WPI; 1997-192904/17.
XX
N-PSDB; AAT91432.
XX
PT New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
PT useful for diagnosis of M. tuberculosis infection.
XX
PS Example 3; Page 138-139; 190pp; English.
XX
CC A new immunogenic polypeptide has been developed comprising an
CC immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
CC variant differing only in conservative substitutions and/or
CC modifications). The present sequence represents a M.tuberculosis antigen,

PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;
 DR WPI; 1998-251292/22.
 DR N-PSDB; AAV44371.

XX New isolated Mycobacterium tuberculosis polypeptides and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and
 PT diagnosis of tuberculosis.

XX Example 3; Page 125-126; 250pp; English.

XX This polypeptide comprises Mycobacterium tuberculosis antigen TbH-9. A
 CC DNA sequence (see AAV44371) coding for antigen TbH-9 was isolated from a
 CC M. tuberculosis strain H37Rv expression library using sera from patients
 CC having pulmonary or pleural tuberculosis. The invention relates to
 CC compositions and methods for diagnosing tuberculosis. It provides
 CC polypeptides (see AAV64291-W64379) comprising an antigenic portion of a
 CC soluble M. tuberculosis antigen, or an immunogenic portion of an M.
 CC tuberculosis antigen, as well as DNA sequences encoding such
 CC polypeptides, recombinant expression vectors and transformed or
 CC transfected host cells. Also claimed are methods and diagnostic kits for
 CC detecting M. tuberculosis infection in a patient using these
 CC polypeptides, antibodies or oligonucleotide probes and primers, for the
 CC diagnosis of tuberculosis. (Updated on 17-OCT-2003 to standardise OS
 CC field)

XX Sequence 263 AA;

Query Match 99.8%; Score 1304; DB 2; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGCAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
 DB 1 VAWMSVTAGCAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
 QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEEAPENTTSAGLLLEQAAAVEEASD 120
 DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEEAPENTTSAGLLLEQAAAVEEASD 120
 QY 121 TAAANQLMNNVPQALKQALQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMSMTN 180
 DB 121 TAAANQLMNNVPQALKQALQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMSMTN 180
 QY 181 SGVSMNTNLTSSMLKGFAPAPAAAQAVQTAQAQNGVRAMSSIGSSLGSGGGVAAANLGRAA 240
 DB 181 SGVSMNTNLTSSMLKGFAPAPAAAQAVQTAQAQNGVRAMSSIGSSLGSGGGVAAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 4
 AAW81680
 ID AAW81680 standard; protein; 263 AA.

XX AAW81680;

XX 27-JAN-1999 (first entry)

XX M. tuberculosis immunogenic polypeptide TbH-9.

XX Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
 XX vaccine; pharmaceutical; infection; diagnosis.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers
 XX Misc-difference 254
 XX /label= unknown

PN W09816646-A2.
 XX 23-APR-1998.
 XX 07-OCT-1997; 97WO-US018293.
 XX 11-OCT-1996; 96US-00730510.
 PR 13-MAR-1997; 97US-00818112.
 XX (CORI-) CORIXA CORP.

XX Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
 PI Vedvick TS, Twardzik DR, Lodes MJ;
 XX WPI; 1998-261042/23.
 DR N-PSDB; AAV64479.

XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and for
 PT diagnosis, treatment and prevention of tuberculosis.

XX Example 3b; Page 119-120; 230pp; English.

XX This sequence represents an immunogenic portion of a soluble
 CC Mycobacterium tuberculosis (MT) antigen which can be used in a method for
 CC inducing protective immunity against tuberculosis (TB). This sequence can
 CC be formulated into vaccines and/or pharmaceutical compositions for
 CC immunising against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis

XX Sequence 263 AA;

Query Match 99.8%; Score 1304; DB 2; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105;
 Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGCAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
 DB 1 VAWMSVTAGCAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
 QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEEAPENTTSAGLLLEQAAAVEEASD 120
 DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEEAPENTTSAGLLLEQAAAVEEASD 120
 QY 121 TAAANQLMNNVPQALKQALQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMSMTN 180
 DB 121 TAAANQLMNNVPQALKQALQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMSMTN 180
 QY 181 SGVSMNTNLTSSMLKGFAPAPAAAQAVQTAQAQNGVRAMSSIGSSLGSGGGVAAANLGRAA 240
 DB 181 SGVSMNTNLTSSMLKGFAPAPAAAQAVQTAQAQNGVRAMSSIGSSLGSGGGVAAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 5
 AAY32062
 ID AAY32062 standard; protein; 263 AA.

XX AAY32062;

XX 17-JAN-2000 (first entry)

XX Mycobacterium tuberculosis antigen TbH9.

XX Tuberculosis; antigen; fusion protein; TbH9; diagnosis; therapy; vaccine;
 XX immunogen.

XX Mycobacterium tuberculosis.

XX Key Location/Qualifiers


```

XX FN WO9942076-A2.
XX PD 26-AUG-1999.
XX PF 17-FEB-1999; 99WO-US003268.
XX PR 18-FEB-1998; 98US-00025197.
XX PR 05-MAY-1998; 98US-00072967.
XX PA (CORI-) CORIXA CORP.
XX PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
XX PI Vedvick TS, Iwardzik DR, Lodes MJ, Hendrickson RC;
XX DR WPI; 1999-527409/44.
XX PT New antigens from Mycobacterium tuberculosis useful in diagnostic skin
XX PT tests and protective or therapeutic vaccines or compositions.
XX PS Example 3; Page 115; 299pp; English.
XX CC The present invention describes polypeptides comprising an immunogenic
XX CC part of a Mycobacterium tuberculosis antigen (Ag). Also described are
XX CC vaccines and fusion protein containing M. tuberculosis Ag's. M.
XX CC tuberculosis Ag's. DNAs encoding them, derived fusion proteins and other
XX CC polypeptides fragments, can be used in pharmaceutical compositions or
XX CC vaccines to generate a protective or therapeutic immune response to M.
XX CC tuberculosis and as reagents in skin tests for diagnosis of tuberculosis.
XX CC Ag can induce proliferation of, or cytokine secretion by, T, B or natural
XX CC killer cells and/or macrophages in tuberculosis-immune subjects. AA219249
XX CC to AA219460 and AA239083 to AA239225 are used in the exemplification of
XX CC the present invention
XX SQ Sequence 263 AA;

Query Match          99.8%; Score 1304; DB 2; Length 263;
Best Local Similarity 100.0%; Pred. No. 3.1e-105;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
DB 1 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60

QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGGLEQAAVEEASD 120
DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLEQAAVEEASD 120

QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVSMMNNHSMNTN 180
DB 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVSMMNNHSMNTN 180

QY 181 SGVSMNTNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240
DB 181 SGVSMNTNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240

QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 8
AAE29706
ID AAE29706 standard; protein; 263 AA.
XX AC AAE29706;
XX DT 27-JAN-2003 (first entry)
XX DE Mycobacterium sp. TbH9 antigenic protein.
XX KW Vaccine; immunity; diagnostic agent; gene therapy; TbH9 antigen.
XX XX

```

```

OS Mycobacterium sp.
XX Key Location/Qualifiers
XX FT Misc-difference 254 /note= "Encoded by NAG"
XX FT
XX PN WO200272792-A2.
XX PD 19-SEP-2002.
XX PF 13-MAR-2002; 2002WO-US008223.
XX PR 13-MAR-2001; 2001US-0275837P.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky Y, Brannon M, Guderian J;
XX PI WPI; 2002-759844/82.
XX DR N-PSDB; AAD47081.
XX PT New recombinant nucleic acid molecule comprising a Leishmania TSA, Leif,
XX PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
XX PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
XX PS tuberculosis.
XX PS Disclosure; Page 84; 155pp; English.
XX CC The invention relates to a recombinant nucleic acid molecule encoding a
XX CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
XX CC polynucleotide sequence encoding an antigen or an antigenic fragment from
XX CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
XX CC polypeptide or its fragment. The Leishmania polynucleotide is selected
XX CC from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention
XX CC are used in methods for eliciting immune response in mammals. They are
XX CC useful as vaccines to elicit protective immunity against pathogenic
XX CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
XX CC polypeptides are used for enhancing the expression of polynucleotides, as
XX CC in vivo diagnostic agents and for raising antibodies in a non-human
XX CC animal. The invention is used in gene therapy. The present sequence is
XX CC Mycobacterium sp. TbH9 antigenic protein
XX SQ Sequence 263 AA;

Query Match          99.8%; Score 1304; DB 5; Length 263;
Best Local Similarity 100.0%; Pred. No. 3.1e-105;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
DB 1 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60

QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGGLEQAAVEEASD 120
DB 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLEQAAVEEASD 120

QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVSMMNNHSMNTN 180
DB 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVSMMNNHSMNTN 180

QY 181 SGVSMNTNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240
DB 181 SGVSMNTNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240

QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
DB 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 9
AAE17570
ID AAE17570 standard; protein; 263 AA.
XX XX

```

AC AAEL1570;
 XX 22-APR-2002 (first entry)
 XX Mycobacterium species MTB39 (TbH9) protein #1.
 DE Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB39; TbH9 protein.
 XX Mycobacterium sp.
 OS
 XX
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 254 /label= Unknown
 FT /note= "Encoded by NAG"
 FN WO200198460-A2.
 XX
 XX
 PD 27-DEC-2001.
 XX
 XX 20-JUN-2001; 2001WO-US019959.
 XX
 XX 20-JUN-2000; 2000US-00597796.
 PR 01-FEB-2001; 2001US-0285737P.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 XX
 XX Skeiky Y, Reed S, Alderson M;
 PI
 XX WPI; 2002-147798/19.
 DR N-PSDB; AAD28340.
 DR
 XX
 XX
 FT Composition comprising MTB39 antigen and MTB32A antigen from
 FT Mycobacterium species, useful for eliciting immune response in a subject.
 XX
 XX
 PS Claim 83; Page 100; 136pp; English.
 XX
 CC The present invention relates to fusion proteins containing at least two
 CC Mycobacterium species antigens, nucleotides encoding them and
 CC compositions comprising such fusion proteins. The present invention
 CC particularly relates to nucleic acids encoding fusion proteins that
 CC include two or more individual M. tuberculosis antigens which increase
 CC the serological sensitivity of sera from individuals infected with
 CC tuberculosis and methods for their use in diagnosis, prevention and
 CC treatment of tuberculosis infection. Sequences of the invention are
 CC useful for eliciting an immune response in a mammal, e.g., human,
 CC immunised with BCG. They are useful in the diagnosis, treatment and
 CC prevention of Mycobacterium infection. The fusion proteins and the
 CC polynucleotides are useful as diagnostic tools in patients infected with
 CC Mycobacterium, in vitro and in vivo assays for detecting humoral
 CC antibodies or cell-mediated immunity against M. tuberculosis, for the
 CC diagnosis of an infection or monitoring of disease progression, as
 CC immunogens to generate or elicit a protective immune response in a
 CC patient and for raising anti-M. tuberculosis antibodies in a non-human
 CC animal. Sequences of the invention are also used as vaccines. MTB32A
 CC fusion proteins of the invention are useful as in vivo diagnostic agents
 CC for intradermal skin test. The present sequence is Mycobacterium species
 CC MTB39 (TbH9) protein
 XX
 SQ Sequence 263 AA;

Query Match 99.8%; Score 1304; DB 5; Length 263;
 Best Local Similarity 100.0%; Pred. No. 3.1e-105; Indels 0; Gaps 0;
 Matches 263; Conservative 0; Mismatches 0;
 QY 1 VAMSVTAGQAEITAAQVRVAAAYATAYGLTVPPVIAENRAELMILITATLLQNTTPA 60
 Db 1 VAMSVTAGQAEITAAQVRVAAAYATAYGLTVPPVIAENRAELMILITATLLQNTTPA 60
 QY 61 IAYNEAYGEMWQAQDAAMFGYAAATATATATATLLPPEAPENTTSAGLLLEQAAAVVEEASD 120
 Db 61 IAYNEAYGEMWQAQDAAMFGYAAATATATATATLLPPEAPENTTSAGLLLEQAAAVVEEASD 120

QY 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPISNNMVMANNHMSMTN 180
 Db 121 TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPISNNMVMANNHMSMTN 180
 QY 181 SGVSMNTLSSMLKGFAPAAAAQAVCTAAQNGVRAMSSLSGLSSGLGGGVAANLGRAA 240
 Db 181 SGVSMNTLSSMLKGFAPAAAAQAVCTAAQNGVRAMSSLSGLSSGLGGGVAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRNGGPA 263
 Db 241 SVRYGHRDGGKYAXSGRNGGPA 263
 RESULT 10
 AAU74591
 ID AAU74591 standard; protein; 358 AA.
 XX
 AC AAU74591;
 XX
 XX 29-AUG-2003 (revised)
 DT 08-MAY-2002 (first entry)
 XX
 DE Antigenic fusion protein TbH9-Tb38-1.
 XX
 KW Fusion protein; tuberculosis; Mycobacterium tuberculosis;
 KW tuberculostatic; immunogen; vaccine; TbH9-Tb38-1; TbH9; Tb38-1.
 XX
 OS Mycobacterium tuberculosis.
 OS Chimeric.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 254 /label= unknown
 FT
 FN US2002009459-A1.
 XX
 PD 24-JAN-2002.
 XX
 XX 07-APR-1999; 99US-00287849.
 PR 13-MAR-1997; 97US-00818112.
 PR 01-OCT-1997; 97US-00342578.
 PR 18-FEB-1998; 98US-00025197.
 PR 07-APR-1998; 98US-00056556.
 PR 30-DEC-1998; 98US-00223040.
 XX
 XX (REED/) REED S G.
 PA (SKEI/) SKEIKY Y A.
 PA (DILL/) DILLON D C.
 PA (ALDE/) ALDERSON M.
 PA (CAMP/) CAMPOS-NETO A.
 XX
 XX Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
 WPI; 2002-171134/22.
 XX
 XX New fusion proteins of Mycobacterium tuberculosis antigens, useful for
 XX diagnosing, treating or preventing M. tuberculosis infection,
 XX particularly as vaccine for treating or preventing tuberculosis.
 XX
 PS Claim 1; Fig 4C-D; 62pp; English.
 XX
 CC The invention relates to a purified polypeptide which induces an immune
 CC response of Mycobacterium tuberculosis. Polypeptides of the invention are
 CC useful for diagnosing, treating or preventing M. tuberculosis infection,
 CC particularly tuberculosis infection. In particular, the polypeptides are
 CC useful as a vaccine formulation with an adjuvant to afford long-term
 CC protection in animals against the development of tuberculosis. The
 CC protein coding sequence may be used to encode a protein product for use
 CC as an immunogen to induce and/or enhance an immune response to M.
 CC tuberculosis. This sequence represents an M. tuberculosis fusion protein
 CC of the invention. Note: The specification states that this polypeptide is
 CC encoded by the polynucleotide shown in ABK14131. (Updated on 29-AUG-2003

CC to standardise OS field)

XX
SQ Sequence 359 AA;

Query Match 99.8%; Score 1304; DB 5; Length 358;
Best Local Similarity 100.0%; Pred. No. 4.7e-105;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAQAELTAAQVRVAAAAYETAYGLTVPPPIAENRAELMILIAITNLLGQNTPA 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1 VAMSVTAQAELTAAQVRVAAAAYETAYGLTVPPPIAENRAELMILIAITNLLGQNTPA 60
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 61 IAVNEAEYGEMWAQDAAMFGYAATAATATATATLTPFEEAPEMTSAGGLEQAAAAVEEASD 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 61 IAVNEAEYGEMWAQDAAMFGYAATAATATATATLTPFEEAPEMTSAGGLEQAAAAVEEASD 120
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 121 TAAANQLMNVPQALKOLAQTQGTTPSSKLGLWKTVPSPHSPISNMVSMANNHMSMTN 180
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 121 TAAANQLMNVPQALKOLAQTQGTTPSSKLGLWKTVPSPHSPISNMVSMANNHMSMTN 180
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 181 SGVSMTNTLSMLKGFAFAAAAQAVQTAQNGVRAMSSLGSSGLGGGVAANLGRAA 240
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 181 SGVSMTNTLSMLKGFAFAAAAQAVQTAQNGVRAMSSLGSSGLGGGVAANLGRAA 240
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 11
AAW32381
ID AAW32381 standard; protein; 391 AA.

AC AAW32381;
XX
DT 13-JAN-1998 (first entry)
DE Mycobacterium tuberculosis antigen TbH-9FL.
XX Antigen; immunogen; vaccine; tuberculosis; non specific adjuvant;
XX skin testing; M.tuberculosis.
KW XX
OS Mycobacterium tuberculosis.
XX
PN WO9709429-A2.
XX
PD 13-MAR-1997.
XX
PF 30-AUG-1996; 96WO-US014675.
PR 01-SEP-1995; 95US-00523435.
PR 22-SEP-1995; 95US-00532136.
PR 22-MAR-1996; 96US-00620280.
PR 05-JUN-1996; 96US-00658800.
PR 12-JUL-1996; 96US-00680573.
XX (CORI-) CORIXA CORP.
XX
PI Reed SG, Skeiky YAW, Dillon DC, Campos-Neto A, Houghton R;
PI Vedwick TH, Twardzik DR;
XX WPI; 1997-192904/17.
DR N-PsDB; AAT91455.
XX New immunogenic polypeptide(s) from soluble M. tuberculosis antigens -
XX useful for diagnosis of M. tuberculosis infection.
PS Example 3; Page 150-152; 190pp; English.
XX

A new immunogenic polypeptide has been developed comprising an
immunogenic part of a soluble Mycobacterium tuberculosis antigen (or its
variant differing only in conservative substitutions and/or
modifications). The present sequence represents a M.tuberculosis antigen.
CC
CC
CC

PT	diagnosis of tuberculosis.
XX	Example 3; Page 133-135; 250pp; English.
CC	This polypeptide comprises Mycobacterium tuberculosis antigen TbH-9FL. It is encoded by genomic DNA (see AAV44395) isolated from a M. tuberculosis strain H37Rv genomic library using a probe from clone TbH-9 (see AAV44371). The invention relates to compositions and methods for diagnosing tuberculosis. It provides polypeptides (see AAW64291-W64379) comprising an antigenic portion of a soluble M. tuberculosis antigen, or an immunogenic portion of an M. tuberculosis antigen, as well as DNA sequences encoding such polypeptides, recombinant expression vectors and transformed or transfected host cells. Also claimed are methods and diagnostic kits for detecting M. tuberculosis infection in a patient using these polypeptides, antibodies or oligonucleotide probes and primers, for the diagnosis of tuberculosis. (Updated on 17-OCT-2003 to standardise OS field)
XX	Sequence 391 AA;
SQ	
	Query Match 90.9%; Score 1187; DB 2; Length 391;
	Best Local Similarity 99.6%; Pred. No. 8.2e-95;
	Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY	1 VAMSVTAGOAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMIATNLGQNTPA 60
DB	74 VAMSVTAGOAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMIATNLGQNTPA 133
OY	61 IAVNEASYGENWQAQDAAMFGYAAATATATATALLPFEAPEMTSAGGLEQAAAAVEEASD 120
DB	134 IAVNEASYGENWQAQDAAMFGYAAATATATATALLPFEAPEMTSAGGLEQAAAAVEEASD 193
OY	121 TAAANQLMNNVPALQOLAQTOGTTPSSKLGKMKTVSPHPSPISNNMVSNANNHMSMTN 180
DB	194 TAAANQLMNNVPALQOLAQTOGTTPSSKLGKMKTVSPHPSPISNNMVSNANNHMSMTN 253
OY	181 SGVSMVTNLTSSMLKGFPAPAAAAQVOTAAQNGVRAMSSIGLSSGLGGGVAANLGSAA 240
DB	254 SGVSMVTNLTSSMLKGFPAPAAAAQVOTAAQNGVRAMSSIGLSSGLGGGVAANLGSAA 313
OY	241 SV 242
DB	314 SV 315
	RESULT 14
AAW81702	
ID	AAW81702 standard; protein; 391 AA.
AC	AAW81702;
XX	27-JAN-1999 (first entry)
XX	M. tuberculosis immunogenic polypeptide TbH-9FL.
DE	Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;
XX	vaccine; pharmaceutical; infection; diagnosis.
KW	

XX WPI; 1998-261042/23.
 DR N-PSDB; AAV64503.
 XX
 XX Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used to
 PT develop products for the detection of M. tuberculosis infection and for
 PT diagnosis, treatment and prevention of tuberculosis.
 XX
 XX Example 3B; Page 128-129; 230pp; English.
 PS
 PS This sequence represents an immunogenic portion of a soluble
 XX Mycobacterium tuberculosis (MT) antigen which can be used in a method for
 CC inducing protective immunity against tuberculosis (TB). This sequence can
 CC be formulated into vaccines and/or pharmaceutical compositions for
 CC immunisation against M. tuberculosis infection or may be used for the
 CC diagnosis of tuberculosis
 XX
 XX Sequence 391 AA;
 SQ

Query Match 90.9%; Score 1187; DB 2; Length 391;
 Best Local Similarity 99.6%; Pred. No. 8.2e-95;
 Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITATNLLGQNTPA 60
 DB 74 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITATNLLGQNTPA 133
 QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEAPEMTSAGLLLEQAAAVEEASD 120
 DB 134 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEAPEMTSAGLLLEQAAAVEEASD 193
 QY 121 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLKTVSPHRSPISNMVMANNHMSMTN 180
 DB 194 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLKTVSPHRSPISNMVMANNHMSMTN 253
 QY 181 SGVSMNTNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSILGSSGLGGGVAANLGRAA 240
 DB 254 SGVSMNTNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSILGSSGLGGGVAANLGRAA 313
 QY 241 SV 242
 DB 314 SV 315

RESULT 15

AY04778
 ID AAY04778 standard; protein; 391 AA.

XX AC AAY04778;

XX DT 06-JUL-1999 (first entry)

XX DE Mycobacterium species protein sequence 5R.

XX KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;
 XX hybridisation; detection; vaccine; immunisation; infection.

XX OS Mycobacterium sp.

XX PN WO9909186-A2.

XX PD 25-FEB-1999.

XX PF 14-AUG-1998; 98WO-FR001813.

XX PR 14-AUG-1997; 97FR-00010404.

XX PR 11-SEP-1997; 97FR-00011325.

XX PA (INSP) INST PASTEUR.

XX PI Gicquel B, Portnoie D, Lim E, Pelicic V, Guigueno A;

XX PI Goguet De La Salmoniere Y;

DR WPI; 1999-181045/15.
 DR N-PSDB; AAX34030.
 XX
 XX Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in infection-associated
 PT protein expression.
 XX
 XX Claim 32; Fig 5R; 309pp; French.
 PS
 PS Sequences AAY04742-Y05000 and AAY07201-Y07204 represent secreted proteins
 CC from various Mycobacterium species microorganisms. The encoding
 CC nucleotide sequences can be used as primers and probes for methods for
 CC detecting and identifying mycobacteria, especially belonging to the M.
 CC tuberculosis complex. The encoded proteins can be used in vaccines for
 CC immunisation against a bacterial or viral infection
 XX
 XX Sequence 391 AA;
 SQ

Query Match 90.9%; Score 1187; DB 2; Length 391;
 Best Local Similarity 99.6%; Pred. No. 8.2e-95;
 Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITATNLLGQNTPA 60
 DB 74 VAWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITATNLLGQNTPA 133
 QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEAPEMTSAGLLLEQAAAVEEASD 120
 DB 134 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEAPEMTSAGLLLEQAAAVEEASD 193
 QY 121 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLKTVSPHRSPISNMVMANNHMSMTN 180
 DB 194 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLKTVSPHRSPISNMVMANNHMSMTN 253
 QY 181 SGVSMNTNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSILGSSGLGGGVAANLGRAA 240
 DB 254 SGVSMNTNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSILGSSGLGGGVAANLGRAA 313
 QY 241 SV 242
 DB 314 SV 315

Search completed: June 30, 2004, 16:48:52

Job time : 28.0437 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:44:57 ; Search time 8.0043 Seconds
(without alignments)
1697.113 Million cell updates/sec

Title: US-09-597-796C-26

Perfect score: 1306

Sequence: 1 VAMSVTAQAEALTAQVRV.....YGRDGGKYXSGRRNGGPA 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:**
1: /cgn2_6/ptodata/2/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/2/iaa/6C COMB.pcp.*
6: /cgn2_6/ptodata/2/iaa/PTUS COMB.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1304	99.8	263	3	US-08-818-112-91
2	1304	99.8	263	4	US-08-818-111-92
3	1304	99.8	263	4	US-09-056-556-91
4	1304	99.8	263	4	US-09-072-596-92
5	1304	99.8	263	4	US-09-072-967-91
6	1304	99.8	358	4	US-09-287-849-8
7	1187	90.9	391	3	US-08-818-112-107
8	1187	90.9	391	4	US-08-818-111-102
9	1187	90.9	391	4	US-09-056-556-107
10	1187	90.9	391	4	US-09-072-596-102
11	1187	90.9	391	4	US-09-072-967-107
12	1187	90.9	596	4	US-09-287-849-26
13	1187	90.9	600	4	US-09-287-849-22
14	1182	90.5	729	4	US-09-223-040-2
15	1182	90.5	729	4	US-09-287-849-2
16	1001.5	76.7	396	3	US-08-818-112-111
17	1001.5	76.7	396	4	US-08-818-111-106
18	1001.5	76.7	396	4	US-09-056-556-111
19	1001.5	76.7	396	4	US-09-072-967-106
20	1001.5	76.7	396	4	US-09-072-967-111
21	949.5	72.7	359	3	US-08-818-112-109
22	949.5	72.7	359	4	US-08-818-111-104
23	949.5	72.7	359	4	US-09-056-556-109
24	949.5	72.7	359	4	US-09-072-596-104
25	949.5	72.7	359	4	US-09-072-967-109
26	429.5	32.9	400	4	US-09-073-009-126
27	429.5	32.9	400	4	US-09-073-010-126

28 325.5 24.9 423 4 US-09-073-009-142 Sequence 142, App
29 325.5 24.9 423 4 US-09-073-010-142 Sequence 142, App
30 325.5 24.9 710 4 US-09-287-849-16 Sequence 16, Appl
31 325.5 24.9 856 4 US-09-287-849-12 Sequence 12, Appl
32 320.5 24.5 943 4 US-09-477-135A-131 Sequence 131, App
33 239 204 4 US-08-311-731A-57 Sequence 57, Appl
34 235 18.0 208 4 US-08-311-731A-208 Sequence 208, App
35 223 17.1 141 4 US-09-073-009-15 Sequence 15, Appl
36 223 17.1 141 4 US-09-073-010-15 Sequence 15, Appl
37 197.5 15.1 943 4 US-09-056-556-204 Sequence 204, App
38 197.5 15.1 943 4 US-09-072-596-199 Sequence 199, App
39 197.5 15.1 943 4 US-09-072-967-204 Sequence 204, App
40 188.5 14.4 368 3 US-08-818-112-114 Sequence 114, App
41 188.5 14.4 368 4 US-08-818-111-109 Sequence 109, App
42 188.5 14.4 368 4 US-09-056-556-114 Sequence 114, App
43 188.5 14.4 368 4 US-09-072-596-109 Sequence 109, App
44 188.5 14.4 368 4 US-09-072-967-114 Sequence 114, App
45 188.5 14.4 371 4 US-09-050-739-92 Sequence 92, Appl

ALIGNMENTS

RESULT 1
US-08-818-112-91
; Sequence 91, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J. 31,392
; REGISTRATION NUMBER: 210121.411C6
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-818-112-91

Query Match 99.8%; Score 1304; DB 3; Length 263;
Best Local Similarity 100.0%; Pred. No. 4.8e-116; Indels 0; Gaps 0;
Matches 263; Conservative 0; Mismatches 0;

Db 121 TAAANQNMNVPOALKQLAQPTGTTTSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180
QY 181 SGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSGLGGVAAANLGRAA 240
Db 181 SGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSGLGGVAAANLGRAA 240
QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 4

US-09-072-596-92
; Sequence 92, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 92:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-596-92

Query Match 99.8%; Score 1304; DB 4; Length 263;
Best Local Similarity 100.0%; Pred. No. 4.8e-116;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPVPIAENRAELMILLIATNLLGQNTPA 60
Db 1 VAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPVPIAENRAELMILLIATNLLGQNTPA 60
QY 61 IAVNEAEYGEWQAQDAAMFGYAAATATATATATLPPFEAPEMTSAGGLLEQAAAVEASD 120
Db 61 IAVNEAEYGEWQAQDAAMFGYAAATATATATLPPFEAPEMTSAGGLLEQAAAVEASD 120
QY 121 TAAANQNMNVPOALKQLAQPTGTTTSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180
Db 121 TAAANQNMNVPOALKQLAQPTGTTTSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180

QY 181 SGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSGLGGVAAANLGRAA 240
Db 181 SGVSMNTLSSMLKGFAPAAAAQAVTAAQNGVRAMSSLGSSGLGGVAAANLGRAA 240
QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 5

US-09-072-967-91
; Sequence 91, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 91:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 263 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-967-91

Query Match 99.8%; Score 1304; DB 4; Length 263;
Best Local Similarity 100.0%; Pred. No. 4.8e-116;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPVPIAENRAELMILLIATNLLGQNTPA 60
Db 1 VAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPVPIAENRAELMILLIATNLLGQNTPA 60
QY 61 IAVNEAEYGEWQAQDAAMFGYAAATATATATLPPFEAPEMTSAGGLLEQAAAVEASD 120
Db 61 IAVNEAEYGEWQAQDAAMFGYAAATATATATLPPFEAPEMTSAGGLLEQAAAVEASD 120
QY 121 TAAANQNMNVPOALKQLAQPTGTTTSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180
Db 121 TAAANQNMNVPOALKQLAQPTGTTTSSKLGGLWKTVPSPHRSPISNMVMANNHSMTN 180

QY 181 SGVSMNTLSSMLKGFAPAAAAAQTAAQNGVRAMSSLSGSSGLGGVAAANLGRAA 240
Db |||||
QY 181 SGVSMNTLSSMLKGFAPAAAAAQTAAQNGVRAMSSLSGSSGLGGVAAANLGRAA 240
Db |||||
QY 241 SVRYGHRDGGKYXSGRRNGGPA 263
Db |||||
QY 241 SVRYGHRDGGKYXSGRRNGGPA 263
Db |||||

RESULT 6

US-09-287-849-8
; Sequence 8, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: Protein Tb9-Tb38-1
; NAME/KEY: MOD_RES
; LOCATION: (254)
; OTHER INFORMATION: Xaa = any amino acid
US-09-287-849-8

Query Match 99.8%; Score 1304; DB 4; Length 358;
Best Local Similarity 100.0%; Pred. No. 7.4e-116;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAMSVTGAQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMILIAIATNLGQNTPA 60
Db 1 VAMSVTGAQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMILIAIATNLGQNTPA 60
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGGLLEQAAVEEASD 120
Db 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGGLLEQAAVEEASD 120
QY 121 TAAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVSPHRSPISNNVSMANNHMSMTN 180
Db 121 TAAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVSPHRSPISNNVSMANNHMSMTN 180
QY 181 SGVSMNTLSSMLKGFAPAAAAAQTAAQNGVRAMSSLSGSSGLGGVAAANLGRAA 240
Db 181 SGVSMNTLSSMLKGFAPAAAAAQTAAQNGVRAMSSLSGSSGLGGVAAANLGRAA 240
QY 241 SVRYGHRDGGKYXSGRRNGGPA 263
Db 241 SVRYGHRDGGKYXSGRRNGGPA 263

RESULT 7

US-08-818-112-107
; Sequence 107, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Twardzik, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-818-112-107

Query Match 90.9%; Score 1187; DB 3; Length 391;
Best Local Similarity 99.6%; Pred. No. 1.1e-104;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAMSVTGAQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMILIAIATNLGQNTPA 60
Db 74 VAMSVTGAQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMILIAIATNLGQNTPA 133
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGGLLEQAAVEEASD 120
Db 134 IAVNEAEYGEWMAQDAAMFGYAAATATATATLTPFEAPEMTSAGGLLEQAAVEEASD 193
QY 121 TAAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVSPHRSPISNNVSMANNHMSMTN 180
Db 194 TAAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVSPHRSPISNNVSMANNHMSMTN 253
QY 181 SGVSMNTLSSMLKGFAPAAAAAQTAAQNGVRAMSSLSGSSGLGGVAAANLGRAA 240
Db 254 SGVSMNTLSSMLKGFAPAAAAAQTAAQNGVRAMSSLSGSSGLGGVAAANLGRAA 313
QY 241 SV 242
Db 314 SV 315

; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 90.9%; Score 1187; DB 4; Length 596;
Best Local Similarity 99.6%; Pred. No. 2e-104;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVPIAENRAELMILIAIATNLLGQNTTPA 60
DB 82 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVPIAENRAELMILIAIATNLLGQNTTPA 141
QY 61 IAVNEAEYGENWQAQDAAMFGYAAATATATATATLLPPEAPEMTSAGGLLEQAAVEEASD 120
DB 142 IAVNEAEYGENWQAQDAAMFGYAAATATATATATLLPPEAPEMTSAGGLLEQAAVEEASD 201
QY 121 TAAANQLMNNVPQALQKLAQPTGGTTPSSKLGGLKTVSPHRSPISNMVSMANNHSMNTN 180
DB 202 TAAANQLMNNVPQALQKLAQPTGGTTPSSKLGGLKTVSPHRSPISNMVSMANNHSMNTN 261
QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSILGSSGLGGGVAANLGRAA 240
DB 262 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSILGSSGLGGGVAANLGRAA 321
QY 241 SV 242
DB 322 SV 323

RESULT 13
US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556

; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

Query Match 90.9%; Score 1187; DB 4; Length 600;
Best Local Similarity 99.6%; Pred. No. 2e-104;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVPIAENRAELMILIAIATNLLGQNTTPA 60
DB 82 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVPIAENRAELMILIAIATNLLGQNTTPA 141
QY 61 IAVNEAEYGENWQAQDAAMFGYAAATATATATLLPPEAPEMTSAGGLLEQAAVEEASD 120
DB 142 IAVNEAEYGENWQAQDAAMFGYAAATATATATLLPPEAPEMTSAGGLLEQAAVEEASD 201
QY 121 TAAANQLMNNVPQALQKLAQPTGGTTPSSKLGGLKTVSPHRSPISNMVSMANNHSMNTN 180
DB 202 TAAANQLMNNVPQALQKLAQPTGGTTPSSKLGGLKTVSPHRSPISNMVSMANNHSMNTN 261
QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSILGSSGLGGGVAANLGRAA 240
DB 262 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSILGSSGLGGGVAANLGRAA 321
QY 241 SV 242
DB 322 SV 323

RESULT 14
US-09-223-040-2
; Sequence 2, Application US/09223040
; Patent No. 6544522
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-223-040-2

Query Match 90.5%; Score 1182; DB 4; Length 729;
Best Local Similarity 99.2%; Pred. No. 8.1e-104;
Matches 240; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVPIAENRAELMILIAIATNLLGQNTTPA 60
DB 215 VAWMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVPIAENRAELMILIAIATNLLGQNTTPA 274
QY 61 IAVNEAEYGENWQAQDAAMFGYAAATATATATLLPPEAPEMTSAGGLLEQAAVEEASD 120
DB 275 IAVNEAEYGENWQAQDAAMFGYAAATATATATLLPPEAPEMTSAGGLLEQAAVEEASD 334

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:52:58 ; Search time 20.5081 Seconds
(without alignments)
3625.462 Million cell updates/sec

Title: US-09-597-796C-26

Perfect score: 1306
Sequence: 1 VAMSVTAGQAEELTAAQVRV.....YGRDGGKXYSGRNGGPA 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1166195 seqs, 282705291 residues

Total number of hits satisfying chosen parameters: 1166195

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA:*

```
1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/prodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/prodata/2/pubpaa/US05_NEW_PUB.pep:*
4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep:*
7: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep:*
8: /cgn2_6/prodata/2/pubpaa/US09_PUBCOMB.pep:*
9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1304	99.8	263	12	Sequence 12, Appl
2	1304	99.8	263	14	Sequence 92, Appl
3	1304	99.8	263	14	Sequence 91, Appl
4	1304	99.8	263	14	Sequence 12, Appl
5	1304	99.8	358	9	Sequence 8, Appl
6	1304	99.8	358	12	Sequence 14, Appl
7	1187	90.9	391	12	Sequence 8, Appl
8	1187	90.9	391	14	Sequence 102, App
9	1187	90.9	391	14	Sequence 107, App
10	1187	90.9	391	14	Sequence 14, Appl
11	1187	90.9	596	12	Sequence 20, Appl
12	1187	90.9	596	12	Sequence 26, Appl
13	1187	90.9	596	14	Sequence 20, Appl
14	1187	90.9	596	14	Sequence 22, Appl
15	1187	90.9	600	9	Sequence 22, Appl

```
16 1187 90.9 600 14 US-10-359-460-22 Sequence 22, Appl
17 1187 90.9 723 15 US-10-369-983-2 Sequence 2, Appl
18 1187 90.9 723 12 US-09-886-349A-18 Sequence 18, Appl
19 1187 90.9 723 14 US-10-098-732A-18 Sequence 18, Appl
20 1187 90.9 723 15 US-10-369-983-21 Sequence 21, Appl
21 1187 90.9 723 15 US-10-369-983-22 Sequence 21, Appl
22 1187 90.9 813 15 US-10-369-983-15 Sequence 15, Appl
23 1187 90.9 825 15 US-10-369-983-14 Sequence 14, Appl
24 1187 90.9 875 15 US-10-369-983-13 Sequence 13, Appl
25 1187 90.9 930 14 US-10-098-732A-65 Sequence 65, Appl
26 1187 90.9 930 15 US-10-369-983-12 Sequence 12, Appl
27 1187 90.9 1010 15 US-10-369-983-4 Sequence 4, Appl
28 1187 90.9 1016 15 US-10-369-983-18 Sequence 18, Appl
29 1187 90.9 1022 15 US-10-369-983-17 Sequence 17, Appl
30 1187 90.9 1154 15 US-10-369-983-16 Sequence 16, Appl
31 1182 90.5 723 9 US-09-287-849-2 Sequence 2, Appl
32 1182 90.5 723 12 US-09-886-349A-16 Sequence 16, Appl
33 1182 90.5 723 14 US-10-359-460-2 Sequence 2, Appl
34 1182 90.5 723 14 US-10-098-732A-16 Sequence 16, Appl
35 1182 90.5 723 15 US-10-359-459-2 Sequence 2, Appl
36 1179 90.3 391 12 US-09-872-186-8 Sequence 8, Appl
37 1001.5 76.7 396 14 US-10-193-003-106 Sequence 106, App
38 1001.5 76.7 396 14 US-10-084-843-111 Sequence 111, App
39 949.5 72.7 359 14 US-10-193-003-104 Sequence 104, App
40 949.5 72.7 359 14 US-10-084-843-109 Sequence 109, App
41 949.5 72.7 393 12 US-10-282-122A-62455 Sequence 62455, A
42 949.5 72.7 393 12 US-10-282-122A-64892 Sequence 64892, A
43 429.5 32.9 400 9 US-09-073-009-126 Sequence 126, App
44 429.5 32.9 400 9 US-09-793-306-126 Sequence 126, App
45 429.5 32.9 405 12 US-10-282-122A-62027 Sequence 62027, A
```

ALIGNMENTS

```
RESULT 1
US-09-886-349A-12
; Sequence 12, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-0090700S
; CURRENT APPLICATION NUMBER: US/09/886.349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 263
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB39 (TbH9)
; FEATURE:
; NAME/KEY: MOD RES
; LOCATION: (254)
; OTHER INFORMATION: Xaa = any amino acid
US-09-886-349A-12
```

Query Match 99.8%; Score 1304; DB 12; Length 263;
Best Local Similarity 100.0%; Pred. No. 1.1e-106;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEELTAAQVRVAAAYETAYGLTVPBPVIAENRAELMILIAIATLLQNTPA 60

DB 1 VAMSVTAGQAEELTAAQVRVAAAYETAYGLTVPBPVIAENRAELMILIAIATLLQNTPA 60

QY 61 IAVNEAEYGEWAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120
 Db 61 IAVNEAEYGEWAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120
 QY 121 TAAANQLMNNVPQALKQLAQPTGGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
 Db 121 TAAANQLMNNVPQALKQLAQPTGGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
 QY 181 SGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSGLSSGLGGVAAANLGRAA 240
 Db 181 SGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSGLSSGLGGVAAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 2

US-10-193-002-92
 ; Sequence 92, Application US/10193002
 ; Publication No. US20030135026A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 Skeiky, Yasir A.W.
 Dillon, Davin C.
 Campos-Neto, Antonia
 Houghton, Raymond
 Vedvick, Thomas S.
 Twardzik, Daniel R.
 Lodes, Michael J.
 Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 350

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/193,002
 FILING DATE: 10-Jul-2002
 CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072,596
 FILING DATE: 05-MAY-1998

ATTORNEY/AGENT INFORMATION:

NAME: Maki, David J.
 REGISTRATION NUMBER: 31,392
 REFERENCE/DOCKET NUMBER: 210121.417C9
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 92:

SEQUENCE CHARACTERISTICS:
 LENGTH: 263 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 92:

US-10-193-002-92

Query Match

Best Local Similarity 99.8%; Score 1304; DB 14; Length 263;

Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEILTAQVRVAAAYETAYGLTVPPVIAENRAELMILITATNLLQONTPA 60
 Db 1 VAMSVTAGQAEILTAQVRVAAAYETAYGLTVPPVIAENRAELMILITATNLLQONTPA 60
 QY 61 IAVNEAEYGEWAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120
 Db 61 IAVNEAEYGEWAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120
 QY 121 TAAANQLMNNVPQALKQLAQPTGGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
 Db 121 TAAANQLMNNVPQALKQLAQPTGGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
 QY 181 SGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSGLSSGLGGVAAANLGRAA 240
 Db 181 SGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRAMSSGLSSGLGGVAAANLGRAA 240
 QY 241 SVRYGHRDGGKYAXSGRRNGGPA 263
 Db 241 SVRYGHRDGGKYAXSGRRNGGPA 263

RESULT 3

US-10-084-843-91

; Sequence 91, Application US/10084843
 ; Publication No. US20030143243A1

GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
 Skeiky, Yasir A.W.
 Dillon, Davin C.
 Campos-Neto, Antonio
 Houghton, Raymond
 Vedvick, Thomas S.
 Twardzik, Daniel R.
 Lodes, Michael J.
 Hendrickson, Ronald C.

TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY AND DIAGNOSIS OF TUBERCULOSIS

NUMBER OF SEQUENCES: 355

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEED and BERRY LLP
 STREET: 6300 Columbia Center, 701 Fifth Avenue
 CITY: Seattle
 STATE: Washington
 COUNTRY: USA
 ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/10/084,843
 FILING DATE: 25-Feb-2002
 CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/072,967
 FILING DATE: 05-MAY-1998
 ATTORNEY/AGENT INFORMATION:
 NAME: Maki, David J.

REGISTRATION NUMBER: 31,392

REFERENCE/DOCKET NUMBER: 210121.411C9

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900
 TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 91:

SEQUENCE CHARACTERISTICS:

LENGTH: 263 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 91:

US-10-084-843-91

Query Match		99.8%; Score 1304; DB 14; Length 263;
Best Local Similarity		100.0%; Pred. No. 1.1e-106;
Matches 263; Conservative		0; Mismatches 0; Indels 0; Gaps 0;
QY	1	VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
DB	1	VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
QY	61	IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120
DB	61	IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120
QY	121	TAAANQLMNNVPOALKOLAQPTGGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
DB	121	TAAANQLMNNVPOALKOLAQPTGGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
QY	181	SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSISGLSSGLGGVAAANLGRAA 240
DB	181	SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSISGLSSGLGGVAAANLGRAA 240
QY	241	SVRYGHRDGGKYAXSGRRNGGPA 263
DB	241	SVRYGHRDGGKYAXSGRRNGGPA 263
RESULT 4		
US-10-098-732A-12		
; Sequence 12, Application US/10098732A		
; Publication No. US20030175294A1		
; GENERAL INFORMATION:		
; APPLICANT: Skeiky, Yasir		
; APPLICANT: Brannon, Mark		
; APPLICANT: Guderian, Jeffrey		
; APPLICANT: Corixa Corporation		
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a		
; FILE REFERENCE: 014058-012010US		
; CURRENT APPLICATION NUMBER: US/10/098,732A		
; CURRENT FILING DATE: 2003-04-29		
; PRIOR APPLICATION NUMBER: US 60/275,837		
; PRIOR FILING DATE: 2001-03-13		
; NUMBER OF SEQ ID NOS: 80		
; SOFTWARE: Patent In Ver. 2.1		
; SEQ ID NO 12		
; LENGTH: 263		
; TYPE: PRT		
; ORGANISM: Mycobacterium tuberculosis		
; FEATURE:		
; OTHER INFORMATION: MTB39 (TbH9)		
; NAME/KEY: MOD RES		
; LOCATION: (254)		
; OTHER INFORMATION: Xaa = any amino acid		
US-10-098-732A-12		
Query Match		99.8%; Score 1304; DB 14; Length 263;
Best Local Similarity		100.0%; Pred. No. 1.1e-106;
Matches 263; Conservative		0; Mismatches 0; Indels 0; Gaps 0;
QY	1	VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
DB	1	VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
QY	61	IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120
DB	61	IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120
QY	121	TAAANQLMNNVPOALKOLAQPTGGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
DB	121	TAAANQLMNNVPOALKOLAQPTGGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
QY	181	SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSISGLSSGLGGVAAANLGRAA 240
Query Match		99.8%; Score 1304; DB 9; Length 358;
Best Local Similarity		100.0%; Pred. No. 1.7e-106;
Matches 263; Conservative		0; Mismatches 0; Indels 0; Gaps 0;
QY	1	VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
DB	1	VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
QY	61	IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120
DB	61	IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120
QY	121	TAAANQLMNNVPOALKOLAQPTGGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
DB	121	TAAANQLMNNVPOALKOLAQPTGGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
QY	181	SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSISGLSSGLGGVAAANLGRAA 240
DB	181	SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSISGLSSGLGGVAAANLGRAA 240
QY	241	SVRYGHRDGGKYAXSGRRNGGPA 263
DB	241	SVRYGHRDGGKYAXSGRRNGGPA 263
RESULT 5		
US-09-287-849-8		
; Sequence 8, Application US/09287849		
; Patent No. US20020009459A1		
; GENERAL INFORMATION:		
; APPLICANT: Reed, Steven G.		
; APPLICANT: Skeiky, Yasir A.W.		
; APPLICANT: Dillon, Davin C.		
; APPLICANT: Alderson, Mark		
; APPLICANT: Campos-Neco, Antonio		
; APPLICANT: Corixa Corporation		
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens		
; TITLE OF INVENTION: and Their Uses		
; FILE REFERENCE: 014058-009020US		
; CURRENT APPLICATION NUMBER: US/09/287,849		
; CURRENT FILING DATE: 1999-04-07		
; PRIOR APPLICATION NUMBER: US 08/818,112		
; PRIOR FILING DATE: 1997-03-13		
; PRIOR APPLICATION NUMBER: US 08/942,578		
; PRIOR FILING DATE: 1997-10-01		
; PRIOR APPLICATION NUMBER: US 09/025,197		
; PRIOR FILING DATE: 1998-02-18		
; PRIOR APPLICATION NUMBER: US 09/056,556		
; PRIOR FILING DATE: 1998-04-07		
; PRIOR APPLICATION NUMBER: US 09/223,040		
; PRIOR FILING DATE: 1998-12-30		
; NUMBER OF SEQ ID NOS: 46		
; SOFTWARE: Patent In Ver. 2.1		
; SEQ ID NO 8		
; LENGTH: 358		
; TYPE: PRT		
; ORGANISM: Artificial Sequence		
; FEATURE:		
; OTHER INFORMATION: Description of Artificial Sequence: bi-fusion		
; OTHER INFORMATION: protein TbH9-Tb38-1		
; NAME/KEY: MOD RES		
; LOCATION: (254)		
; OTHER INFORMATION: Xaa = any amino acid		
US-09-287-849-8		
Query Match		99.8%; Score 1304; DB 9; Length 358;
Best Local Similarity		100.0%; Pred. No. 1.7e-106;
Matches 263; Conservative		0; Mismatches 0; Indels 0; Gaps 0;
QY	1	VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
DB	1	VAMSVTAGQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
QY	61	IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120
DB	61	IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGGLLEQAAAVEEASD 120
QY	121	TAAANQLMNNVPOALKOLAQPTGGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
DB	121	TAAANQLMNNVPOALKOLAQPTGGTTPSSKLGGLWKTVPSPHSPISNMVMANNHSMNTN 180
QY	181	SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSISGLSSGLGGVAAANLGRAA 240
DB	181	SGVSMNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSISGLSSGLGGVAAANLGRAA 240
QY	241	SVRYGHRDGGKYAXSGRRNGGPA 263
DB	241	SVRYGHRDGGKYAXSGRRNGGPA 263

```
RESULT 6
US-10-359-460-8
; Sequence 8, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-0090700S
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: bi-fusion
; OTHER INFORMATION: protein TBH9-Tb38-1
; NAME/KEY: MOD RES
; LOCATION: (254)
; OTHER INFORMATION: Xaa = any amino acid
US-10-359-460-8

Query Match 99.8%; Score 1304; DB 14; Length 358;
Best Local Similarity 100.0%; Pred. No. 1.7e-106;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
DB 1 VAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATATLLPFEAPEMTSAGGLEQAAAVEASD 120
DB 61 IAVNEAEYGENWAQDAAMFGYAAATATATATATLLPFEAPEMTSAGGLEQAAAVEASD 120
QY 121 TAAANQLMNVPAQLKQLAQTQGTTPSSKLGGLWKTVPSPISNNVSMANNHMTN 180
DB 121 TAAANQLMNVPAQLKQLAQTQGTTPSSKLGGLWKTVPSPISNNVSMANNHMTN 180
QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSGLSSGLGGGVAANLGRAA 240
DB 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSGLSSGLGGGVAANLGRAA 240
QY 241 SVRYGHRDGGKYXSGRRNGCPA 263
DB 241 SVRYGHRDGGKYXSGRRNGCPA 263

RESULT 7
US-09-886-349A-14
; Sequence 14, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-0090700S
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 8
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: bi-fusion
; OTHER INFORMATION: protein TBH9-Tb38-1
; NAME/KEY: MOD RES
; LOCATION: (254)
; OTHER INFORMATION: Xaa = any amino acid
US-10-359-460-8

Query Match 99.8%; Score 1304; DB 14; Length 358;
Best Local Similarity 100.0%; Pred. No. 1.7e-106;
Matches 263; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
DB 1 VAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATATLLPFEAPEMTSAGGLEQAAAVEASD 120
DB 61 IAVNEAEYGENWAQDAAMFGYAAATATATATATLLPFEAPEMTSAGGLEQAAAVEASD 120
QY 121 TAAANQLMNVPAQLKQLAQTQGTTPSSKLGGLWKTVPSPISNNVSMANNHMTN 180
DB 121 TAAANQLMNVPAQLKQLAQTQGTTPSSKLGGLWKTVPSPISNNVSMANNHMTN 180
QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSGLSSGLGGGVAANLGRAA 240
DB 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSGLSSGLGGGVAANLGRAA 240
QY 241 SVRYGHRDGGKYXSGRRNGCPA 263
DB 241 SVRYGHRDGGKYXSGRRNGCPA 263
```

```
RESULT 8
US-10-193-002-102
; Sequence 102, Application US/10193002
; Publication No. US20030135026A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-0090700S
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB39 (TbH9FL)
; US-09-886-349A-14

Query Match 90.9%; Score 1187; DB 12; Length 391;
Best Local Similarity 99.6%; Pred. No. 3.8e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 60
DB 74 VAMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIAIATNLLGQNTPA 133
QY 61 IAVNEAEYGENWAQDAAMFGYAAATATATATATLLPFEAPEMTSAGGLEQAAAVEASD 120
DB 134 IAVNEAEYGENWAQDAAMFGYAAATATATATATLLPFEAPEMTSAGGLEQAAAVEASD 193
QY 121 TAAANQLMNVPAQLKQLAQTQGTTPSSKLGGLWKTVPSPISNNVSMANNHMTN 180
DB 194 TAAANQLMNVPAQLKQLAQTQGTTPSSKLGGLWKTVPSPISNNVSMANNHMTN 253
QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSGLSSGLGGGVAANLGRAA 240
DB 254 SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSGLSSGLGGGVAANLGRAA 313
QY 241 SV 242
DB 314 SV 315
```

; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/193,002
 ; FILING DATE: 10-Jul-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072,596
 ; FILING DATE: 05-MAY-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.417C9
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 102:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 391 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 102:
 ; US-10-193-002-102

Query Match 90.9%; Score 1187; DB 14; Length 391;
 Best Local Similarity 99.6%; Pred. No. 3.8e-96;
 Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITNLLGQNTPA 60
 DB 74 VAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITNLLGQNTPA 133
 QY 61 IAVNEAYEGEMWAQDAAMFGVAAAATATATATLLPPEEAPEMTSAGGLLEQAAAVEEASD 120
 DB 134 IAVNEAYEGEMWAQDAAMFGVAAAATATATATLLPPEEAPEMTSAGGLLEQAAAVEEASD 193
 QY 121 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMTN 180
 DB 194 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMTN 253
 QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSGLGGGVAANLGRAA 240
 DB 254 SGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSGLGGGVAANLGRAA 313
 QY 241 SV 242
 DB 314 SV 315

RESULT 9
 US-10-084-843-107
 ; Sequence 107, Application US/10084843
 ; Publication No. US20030143243A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; Skeiky, Yasir A.W.
 ; Dillon, David C.
 ; Campos-Neto, Antonio
 ; Houghton, Raymond
 ; Vedwick, Thomas S.
 ; Twardzik, Daniel R.
 ; Lodes, Michael J.
 ; Hendrickson, Ronald C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 ; AND DIAGNOSIS OF TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 355
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington

; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/10/084,843
 ; FILING DATE: 25-Feb-2002
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072,967
 ; FILING DATE: 05-MAY-1998
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.411C9
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 107:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 391 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 107:
 ; US-10-084-843-107

Query Match 90.9%; Score 1187; DB 14; Length 391;
 Best Local Similarity 99.6%; Pred. No. 3.8e-96;
 Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITNLLGQNTPA 60
 DB 74 VAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMILITNLLGQNTPA 133
 QY 61 IAVNEAYEGEMWAQDAAMFGVAAAATATATATLLPPEEAPEMTSAGGLLEQAAAVEEASD 120
 DB 134 IAVNEAYEGEMWAQDAAMFGVAAAATATATATLLPPEEAPEMTSAGGLLEQAAAVEEASD 193
 QY 121 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMTN 180
 DB 194 TAAANQLMNNVPQALQKLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMTN 253
 QY 181 SGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSGLGGGVAANLGRAA 240
 DB 254 SGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSGLGGGVAANLGRAA 313
 QY 241 SV 242
 DB 314 SV 315

RESULT 10
 US-10-098-732A-14
 ; Sequence 14, Application US/10098732A
 ; Publication No. US20030175294A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; Applicant: Brannon, Mark
 ; Applicant: Guderian, Jeffrey
 ; Applicant: Corixa Corporation
 ; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
 ; FILE OF INVENTION: Leishmania Antigen
 ; FILE REFERENCE: 014058-012010US
 ; CURRENT APPLICATION NUMBER: US/10/098,732A
 ; CURRENT FILING DATE: 2003-04-29
 ; PRIOR APPLICATION NUMBER: US 60/275,837
 ; PRIOR FILING DATE: 2001-03-13
 ; NUMBER OF SEQ ID NOS: 80
 ; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 14
; LENGTH: 391
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: MTB39 full length (TBH9FL)
US-10-038-732A-14

Query Match 90.9%; Score 1187; DB 14; Length 391;
Best Local Similarity 99.6%; Pred. No. 3.6e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	VAMSVTAGQAE	LTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA	60
DB	74	VAMSVTAGQAE	LTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA	133
QY	61	IAVNEAEYGE	MAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD	120
DB	134	IAVNEAEYGE	MAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD	193
QY	121	TAAANQLMNN	VPQALKOLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN	180
DB	194	TAAANQLMNN	VPQALKOLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN	253
QY	181	SGVSMNTN	TLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSGLGGVAAANLGRAA	240
DB	254	SGVSMNTN	TLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSGLGGVAAANLGRAA	313
QY	241	SV	242	
DB	314	SV	315	

RESULT 11
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 90.9%; Score 1187; DB 9; Length 596;
Best Local Similarity 99.6%; Pred. No. 6.6e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAE LTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
DB 82 VAMSVTAGQAE LTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 141
QY 61 IAVNEAEYGE MAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120
DB 142 IAVNEAEYGE MAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 201
QY 121 TAAANQLMNN VPQALKOLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 180
DB 202 TAAANQLMNN VPQALKOLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 261
QY 181 SGVSMNTN TLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240
DB 262 SGVSMNTN TLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 321
QY 241 SV 242
DB 322 SV 323

RESULT 12
US-09-886-349A-20
; Sequence 20, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-886-349A-20

Query Match 90.9%; Score 1187; DB 12; Length 596;
Best Local Similarity 99.6%; Pred. No. 6.6e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAMSVTAGQAE LTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 60
DB 82 VAMSVTAGQAE LTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGQNTPA 141
QY 61 IAVNEAEYGE MAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120
DB 142 IAVNEAEYGE MAQDAAMFGYAAATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 201
QY 121 TAAANQLMNN VPQALKOLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 180
DB 202 TAAANQLMNN VPQALKOLAQPTQGTTPSSKLGGLWKTVPSPISNMVSMANNHSMTN 261
QY 181 SGVSMNTN TLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 240
DB 262 SGVSMNTN TLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGLSSGLGGVAAANLGRAA 321
QY 241 SV 242
DB 322 SV 323

RESULT 13
US-10-359-460-26
; Sequence 26, Application US/10359460
; Publication No. US2003014791A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-012010US
; CURRENT FILING DATE: 2003-04-29
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-10-359-460-26

Query Match 90.9%; Score 1187; DB 14; Length 596;
Best Local Similarity 99.6%; Pred. No. 6.6e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	VAMSVTAGQAEELTAQVRAAAAYETAYGLTVPVIAENRAELMILIAIATNLGQNTPA	60
DB	82	VAMSVTAGQAEELTAQVRAAAAYETAYGLTVPVIAENRAELMILIAIATNLGQNTPA	141
QY	61	IAVNEAEYGEWQAQDAAMFGYAAATATATATATLLPFEEAPMTSAGGLEQAAAVEASD	120
DB	142	IAVNEAEYGEWQAQDAAMFGYAAATATATATLLPFEEAPMTSAGGLEQAAAVEASD	201
QY	121	TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMTN	180
DB	202	TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMTN	261
QY	181	SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSGLGGGVAANLGRAA	240
DB	262	SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSGLGGGVAANLGRAA	321
QY	241	SV 242	
DB	322	SV 323	

RESULT 14
US-10-098-732A-20
; Sequence 20, Application US/10098732A
; Publication No. US2003017529A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeffrey
; APPLICANT: Corixa Corporation

Query Match 90.9%; Score 1187; DB 14; Length 596;
Best Local Similarity 99.6%; Pred. No. 6.6e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	VAMSVTAGQAEELTAQVRAAAAYETAYGLTVPVIAENRAELMILIAIATNLGQNTPA	60
DB	82	VAMSVTAGQAEELTAQVRAAAAYETAYGLTVPVIAENRAELMILIAIATNLGQNTPA	141
QY	61	IAVNEAEYGEWQAQDAAMFGYAAATATATATLLPFEEAPMTSAGGLEQAAAVEASD	120
DB	142	IAVNEAEYGEWQAQDAAMFGYAAATATATATLLPFEEAPMTSAGGLEQAAAVEASD	201
QY	121	TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMTN	180
DB	202	TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMTN	261
QY	181	SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSGLGGGVAANLGRAA	240
DB	262	SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSGLGGGVAANLGRAA	321
QY	241	SV 242	
DB	322	SV 323	

; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 20
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: Protein TbH9-Ra35 (designated MTB59f)
US-10-098-732A-20

Query Match 90.9%; Score 1187; DB 14; Length 596;
Best Local Similarity 99.6%; Pred. No. 6.6e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY	1	VAMSVTAGQAEELTAQVRAAAAYETAYGLTVPVIAENRAELMILIAIATNLGQNTPA	60
DB	82	VAMSVTAGQAEELTAQVRAAAAYETAYGLTVPVIAENRAELMILIAIATNLGQNTPA	141
QY	61	IAVNEAEYGEWQAQDAAMFGYAAATATATATLLPFEEAPMTSAGGLEQAAAVEASD	120
DB	142	IAVNEAEYGEWQAQDAAMFGYAAATATATATLLPFEEAPMTSAGGLEQAAAVEASD	201
QY	121	TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMTN	180
DB	202	TAAANQLMNNVPQALKQLAQPTQGTTPSSKLGGLWKTVPSPHSPISNMVMANNHMTN	261
QY	181	SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSGLGGGVAANLGRAA	240
DB	262	SGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLSGSSGLGGGVAANLGRAA	321
QY	241	SV 242	
DB	322	SV 323	

RESULT 15
US-09-287-849-22
; Sequence 22, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600

```
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-22

Query Match          90.9%; Score 1187; DB 9; Length 600;
Best Local Similarity 99.6%; Pred. No. 6.7e-96;
Matches 241; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VAWMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPPVIAENRAELMILIAATNLLGQNTPA 60
Db 82 VAWMSVTAGQAEHTAAQVRVAAAAYETAYGLTVPPPVIAENRAELMILIAATNLLGQNTPA 141

QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 120
Db 142 IAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGGLLEQAAVEEASD 201

QY 121 TAAANQLMNNVPOALQOLAQPTGGTTPSSKLGGLMKTVPSPHRSPISNMVMANNHMSMTN 180
Db 202 TAAANQLMNNVPOALQOLAQPTGGTTPSSKLGGLMKTVPSPHRSPISNMVMANNHMSMTN 261

QY 181 SGVSMNTNLTSSMLKGFAPAAAQAVCTAAQNGVRAMSSLGSSGLGGGVAANLGRAA 240
Db 262 SGVSMNTNLTSSMLKGFAPAAAQAVCTAAQNGVRAMSSLGSSGLGGGVAANLGRAA 321

QY 241 SV 242
Db 322 SV 323
```

Search completed: June 30, 2004, 17:14:49
Job time : 21.5081 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:31 ; Search time 6.87361 Seconds
(without alignments)
3680.509 Million cell updates/sec

Title: US-09-597-796C-26
Perfect score: 1306
Sequence: 1 VAMSVTAGQAEITAAQVRV.....YGHRDGKXAXSGRRNGGPA 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 78:*
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1187	90.9	391	2 B70608	probable PPE prote
2	1001.5	76.7	396	2 H70741	probable PPE prote
3	949.5	72.7	393	2 C70568	probable PPE prote
4	460	35.2	403	2 H70931	probable PPE prote
5	441	33.8	423	2 B70931	probable PPE prote
6	432.5	33.1	393	2 G70929	probable PPE prote
7	424.5	32.5	421	2 H87056	PPE-family protein
8	422	32.3	409	2 A70932	probable PPE prote
9	403	30.9	413	2 F70560	probable PPE prote
10	402	30.8	391	2 B70625	probable PPE prote
11	402	30.8	463	2 C70931	probable PPE prote
12	395	30.2	408	2 G70925	probable PPE prote
13	389	29.8	394	2 G70881	probable PPE prote
14	386.5	29.6	468	2 A70932	probable PPE prote
15	379	29.0	380	2 A70646	probable PPE prote
16	375	28.7	385	2 H70503	probable PPE prote
17	352	27.0	402	2 A70882	probable PPE prote
18	348.5	26.7	365	2 G70929	probable PPE prote
19	346.5	26.5	394	2 A70504	probable PPE prote
20	346	26.5	350	2 H70929	probable PPE prote
21	339	25.2	391	2 D70922	probable PPE prote
22	328	25.1	406	2 E70675	probable PPE prote
23	325.5	24.9	423	2 C70582	probable PPE prote
24	320.5	24.5	3300	2 D70575	probable PPE prote
25	317	24.3	391	2 A70663	probable PPE prote
26	282	21.6	963	2 B70524	probable PPE prote
27	272	20.8	580	2 G70570	probable PPE prote
28	271.5	20.8	3716	2 E70969	probable PPE prote
29	266	20.4	180	2 G70834	probable PPE prote

30	257	19.7	618	2 H70552	probable PPE prote
31	255.5	19.6	678	2 A70762	probable PPE prote
32	250.5	19.2	1053	2 E70987	probable PPE prote
33	245	18.8	615	2 E70663	probable PPE prote
34	243.5	18.6	443	2 C70780	probable PPE prote
35	243	18.6	487	2 C70830	probable PPE prote
36	242.5	18.6	346	2 H70874	probable PPE prote
37	242	18.5	2523	2 F70846	probable PPE prote
38	239	18.3	655	2 A70931	probable PPE prote
39	238	18.2	3157	2 E70969	probable PPE prote
40	236	18.1	479	2 D70676	probable PPE prote
41	235	18.0	204	2 T45168	probable PPE prote
42	231.5	17.7	987	2 E70808	probable PPE prote
43	229	17.5	1436	2 B70520	probable PPE prote
44	227	17.4	582	2 F70675	probable PPE prote
45	223.5	17.1	645	2 F70825	probable PPE prote

ALIGNMENTS

RESULT 1

B70608
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70608
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: B70608
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-391 <COL>
A:Cross-references: GB:Z93777; GB:AL123456; NID:g3261726; PIDN:CAB07839.1; PID:e311073; I
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match	90.9%	Score 1187;	DB 2;	Length 391;
Best Local Similarity	99.6%	Pred. No. 8.6e-74;		
Matches 241;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	VAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMILIAITNLLGQNTPA	60	
Db	74	VAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMILIAITNLLGQNTPA	133	
Qy	61	I AVNEAYGENWAQDAAMFGYAAATATATATLIPFEAPEMTSAGLLFOAAVEEASD	120	
Db	134	I AVNEAYGENWAQDAAMFGYAAATATATATLIPFEAPEMTSAGLLFOAAVEEASD	193	
Qy	121	TAAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVSPhrPISNNVSMANNHMTN	180	
Db	194	TAAANQLMNNVPQALKOLAQPTQGTTPSSKLGGLWKTVSPhrPISNNVSMANNHMTN	253	
Qy	181	SGVSMNTLSSMLKGFAPAAAAQAVOTAAQNGVRAMSSSLGSSGLGGVAAANLGRAA	240	
Db	254	SGVSMNTLSSMLKGFAPAAAAQAVOTAAQNGVRAMSSSLGSSGLGGVAAANLGRAA	313	
Qy	241	SV 242		
Db	314	SV 315		

RESULT 2

H70741
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: H70741
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: H70741
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-396 <COL>
A: Cross-references: GB:275555; GB:AL123456; NID:G3261608; PIDN:CAA99966.1; PID: e250360;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: PPE

Query Match 76.7%; Score 1001.5; DB 2; Length 396;
Best Local Similarity 84.1%; Pred. No. 3.9e-61;
Matches 207; Conservative 13; Mismatches 21; Indels 5; Gaps 2;

QY 1 VAMSVTAGOELTAAQVRAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGONTTPA 60

DB 74 VAMSVTAGOELTAAQVRAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGONTTPA 133

QY 61 IAVNEAEYGENWQAADAAAMFGYAAATATATATLLPPEEAPMTSAGGLLEQAAVVEASD 120

DB 134 IAVNEAEYGENWQAADAAAMFGYAAATATATALLPPEEAPMTSAGGLLEQAAVVEASD 193

QY 121 TAAANQLMNNVPOALQKLAQTOGTTPSSKLGGLWKTVPSPHRSPIKNVSMANNHSMNTN 180

DB 194 TAAANQLMNNVPOALQKLAQTOGTTPSSKLGGLWKTVPSPHRSPIKNVSMANNHSMNTN 253

QY 181 SGVSMNTLSSMLKGFAPAAAQAQVOTAAONGVAMSS---LGSSLGSSGLGGVAAANL 236

DB 254 SGVSMNTLSSMLKGFAPAAAQAQVOTAAONGVAMSS---LGSSLGSSGLGGVAAANL 312

QY 237 GRAASV 242

DB 313 GRAASV 318

RESULT 3
C70568
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C: Species: Mycobacterium tuberculosis
C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C: Accession: C70568
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: C70568
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-393 <COL>
A: Cross-references: GB:295390; GB:AL123456; NID:G3261766; PIDN: CAB08702.1; PID: e316074;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: PPE

Query Match 72.7%; Score 949.5; DB 2; Length 393;
Best Local Similarity 79.7%; Pred. No. 1.4e-57;
Matches 196; Conservative 14; Mismatches 31; Indels 5; Gaps 2;

QY 1 VAMSVTAGOELTAAQVRAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGONTTPA 60

DB 74 VAMSVTAGOELTAAQVRAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGONTTPA 133

QY 61 IAVNEAEYGENWQAADAAAMFGYAAATATATATLLPPEEAPMTSAGGLLEQAAVVEASD 120

DB 134 IEANQAYSNWQDAAEMGYAATAATATATALLPPEEAPMTSAGGLLEQAAVVEASD 193

QY 121 TAAANQLMNNVPOALQKLAQTOGTTPSSKLGGLWKTVPSPHRSPIKNVSMANNHSMNTN 180

DB 194 TAAANQLMNNVPOALQKLAQTOGTTPSSKLGGLWKTVPSPHRSPIKNVSMANNHSMNTN 253

QY 181 SGVSMNTLSSMLKGFAPAAAQAQVOTAAONGVAMSS---LGSSLGSSGLGGVAAANL 236

DB 254 TGVSMNTLSSMLKGLAPAAAQAQVOTAAONGVAMSS---LGSSLGSSGLGGVAAANL 312

QY 237 GRAASV 242

DB 313 GRAASV 318

RESULT 4
H70931
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C: Species: Mycobacterium tuberculosis
C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C: Accession: H70931
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: H70931
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-403 <COL>
A: Cross-references: GB:AL022021; GB:AL123456; NID:G3250699; PIDN: CAA17728.1; PID: e125461;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: PPE

Query Match 35.2%; Score 460; DB 2; Length 403;
Best Local Similarity 43.3%; Pred. No. 3.4e-24;
Matches 117; Conservative 36; Mismatches 73; Indels 44; Gaps 8;

QY 1 VAMSVTAGOELTAAQVRAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGONTTPA 60

DB 77 VAMSVTAGOELTAAQVRAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGONTTPA 136

QY 61 IAVNEAEYGENWQAADAAAMFGYAAATATATATLLPPEEAPMTSAGGLLEQAAVVEASD 120

DB 137 IAVNEAEYGENWQAADAAAMFGYAAATATATLLPPEEAPMTSAGGLLEQAAVVEASD 195

QY 121 TAAANQLMNNVPOALQKLAQTOGTTPSSKLGGLWKTVPSPHRSPIKNVSMANNHSMNTN 161

DB 196 ASAGAQQTTLSQLIAAIPSVLQGLSSSTAAATFASGGLLIGVSGSSWLDKMLWALDPN 255

QY 162 RSPISNMVSMANNHSMNTNNGVSM-TNTLS---SNLKGFAFAAAAQAQVOTAAONGVAMSS 217

DB 256 -----SNFWNTIASSGLFIPSNITAFPLGLGGVAAADAAGDVLGEATSG----- 300

QY 218 SIGSSL-----GSSGLGGVAAANGRAASV 242

DB 301 GLGGALVAPLGSAGGLGGTVAAGLGNAAV 330

RESULT 5
B70931
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C: Species: Mycobacterium tuberculosis
C: Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C: Accession: B70931
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998

A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: B70931
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-423 <COL>
A:Cross-references: GB:AL020201; GB:AL123456; NID:g3250699; PIDN:CAA1772.1; PID:e125460
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 33.1%; Score 432.5; DB 2; Length 393;
Best Local Similarity 41.0%; Pred. No. 2.4e-22;
Matches 107; Conservative 38; Mismatches 85; Indels 31; Gaps 7;
C:Species: Mycobacterium tuberculosis (strain H37RV)
C:Date: 17-Jul-1998 #sequence_revision 38; text_change 22-Oct-1999
C:Accession: G70929
R;Cole, S.T.; Broch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70929
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-393 <COL>
A:Cross-references: GB:AL020201; GB:AL123456; NID:g3250699; PIDN:CAA1771.1; PID:e125460
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 33.1%; Score 432.5; DB 2; Length 393;
Best Local Similarity 41.0%; Pred. No. 2.4e-22;
Matches 107; Conservative 38; Mismatches 85; Indels 31; Gaps 7;
C:Species: Mycobacterium tuberculosis (strain H37RV)
C:Date: 17-Jul-1998 #sequence_revision 38; text_change 22-Oct-1999
C:Accession: G70929
R;Cole, S.T.; Broch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70929
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-393 <COL>
A:Cross-references: GB:AL020201; GB:AL123456; NID:g3250699; PIDN:CAA1771.1; PID:e125460
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 33.1%; Score 432.5; DB 2; Length 393;
Best Local Similarity 41.0%; Pred. No. 2.4e-22;
Matches 107; Conservative 38; Mismatches 85; Indels 31; Gaps 7;
C:Species: Mycobacterium tuberculosis (strain H37RV)
C:Date: 17-Jul-1998 #sequence_revision 38; text_change 22-Oct-1999
C:Accession: G70929
R;Cole, S.T.; Broch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70929
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-393 <COL>
A:Cross-references: GB:AL020201; GB:AL123456; NID:g3250699; PIDN:CAA1771.1; PID:e125460
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Db 192 TAQSTLTEMITGLPNAQLQSTLPSLQSS-NGPLSLWMLQILFGTFNFTSALLTDIQPY 250
QY 162 RSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSIGS 221
Db 251 ASFPYNTTEGLPYFSGNNGNFIQSAKTL-GLIGSAAAPAAVA-----AAGDAKGLPGLGG 304
QY 222 SLGSSGLGGGVAANLGRAASV 242
Db 305 MLG-----GGPVAAGLGRAASV 321

RESULT 7
H87056
PPE-family protein [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: H87056
R;Cole, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, S.; Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: H87056
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-421 <STO>
A:Cross-references: GB:AL450380; NID:g13093150; PIDN:CAC31563.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML1182

Query Match 32.5%; Score 424.5; DB 2; Length 421;
Best Local Similarity 39.8%; Pred. No. 9.3e-22;
Matches 104; Conservative 36; Mismatches 99; Indels 29; Gaps 5;
C:Species: Mycobacterium tuberculosis (strain H37RV)
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70932
R;Cole, S.T.; Broch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70932
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

Query Match 32.5%; Score 424.5; DB 2; Length 421;
Best Local Similarity 39.8%; Pred. No. 9.3e-22;
Matches 104; Conservative 36; Mismatches 99; Indels 29; Gaps 5;
C:Species: Mycobacterium tuberculosis (strain H37RV)
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70932
R;Cole, S.T.; Broch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70932
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA

Query Match 30.8%; Score 402; DB 2; Length 463;
Best Local Similarity 35.7%; Pred.No.3.6e-20;
Matches 114; Conservative 37; Mismatches 106; Indels 62; Gaps 8;

Qy 1 VAMSVTAGOAEELTAAQVRVAAAAYETAYGLTVPPPIAENRAEMLILIAATNLGQNTPA 60
| | | | |
Db 72 VAWMSATAALAREAAAAQAASAAAAAYEAAPATVPFPVVANRAELAVALIATNIFGQTGA 131

Qy 61 IAVNEASYGENWAOADAAMEFYAAATATATATLLPPEAPEMTSAGLLEQAAAVEASD 120
| | | | |
Db 132 IAFAEARYAEMWAOADAAAMYGYAGSSVSAT--QVTFFAAPPTTTNAAGLATGGVAOAQG 190

Qy 121 TAAANQLMNNYPOALKOLAOPGTQCTPSKLGGLWKTVS--PHRSPI-----SN 167
| | | | |
Db 191 ASAGN--ARSLVSEVLFEIA--TAGTNYNKVIVASLMNAVTCVPYASSVNMLGLGFRESK 247

Qy 168 MVSMANNHM-----SMTNSGVSMNTLTLSMKLKGFPAPA 200
| | | : : : : :
Db 248 MVLPAANDTVITFGMVQFOKFNPVTFPNDFLPKLGALGAGLGRSAISGLGSTAPAI 307

Qy 201 AAQAVQTAQAQNVR-----AMSLSGLSGSLGGGVAANLCRA-----ASV 242
| | | : : : : :
Db 308 SAGASQAGSVCGMSVPSWAATAATPATRTVAAFVSTGLQAVPAATISEGILLSOMALLASV 367

Qy 243 RYGHDRDGKYAXSGRRNGG 261
| | | : : : : :
Db 368 AGGALGGAARAATGCGFLGG 386

RESULT 12
G70925
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C/Accession: G70925
R/Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A>Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; UID:98295987; PMID:9634230
A/Accession: G70925
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-408 <COL>
A/Cross-references: GS:274024; GB:AL123456; NID:g3250700; PIDN:CAB98377.1; PID:el301025;
A/Experimental source: strain H37rv
C/Genetics:
A/Gene: PPE

Query Match 30.2%; Score 395; DB 2; Length 408;
Best Local Similarity 38.9%; Pred.No.9.2e-20;
Matches 110; Conservative 31; Mismatches 78; Indels 64; Gaps 11;

Qy 1 VAMSVTAGOAEELTAAQVRVAAAAYETAYGLTVPPPIAENRAEMLILIAATNLGQNTPA 60
| | | | |
Db 73 VAWLSATAGQAEQAGMQARAAAAAYELAFAMTVPPPWWANRALVALVATNFPGQNTPA 132

Qy 61 IAVNEASYGENWAOADAAMEFYAAATATATATLLPPEAPEMTSAGLLEQAAA-----V 115
| | | | |
Db 133 IAATEAQTAEMWAOADAAMTYAGAIAIAT--ELTPFTAPVTSIPALAQAAATVSVTV 191

Qy 116 EASDTAAANQLMNN-----VP--QALKQ-LAQFTQGTTPSKL-----GGL 154
| | | : : : : :
Db 192 PLLATTAAVPLQLLOLSSTSLSIPWSALQQWLAEANLLGLTPDNRMVTIRLLIGISYFDEGL 251

Qy 155 WKTVSPHRSPIISMVMNMNHMTNSGVSMNTLTLSMKLKGFPAAAAQAVQTAQNVR 214
| | | : : : : :
Db 252 -----LQPFASLAQAQPOTPGAG--DSGSVLDWSWGPTIFA-----GPR 290

Qy 215 AMSSL--GSSLGS-----SGLGGGVAANLCRAAS 241

Db 291 ASPSVAGGAGVGGVQTPOPYWYWALDRESIGSVSAAALGKSS 333

RESULT 13

G70881

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C/Accession: G70881

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A/Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: G70881

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-394 <COL>

A/Cross-references: GB:AL008967; GB:AL123456; NID:93261491; PIDN:CAA15564.1; PID:ell173899

A/Experimental source: strain H37RV

C/Genetics:

A/Gene: PPE

Query Match 29.8%; Score 389; DB 2; Length 394;

Best Local Similarity 37.6%; Pred. No. 2.3e-19;

Matches 103; Conservative 43; Mismatches 106; Indels 22; Gaps 7;

Qy 1 VAMSVTAGQAEATAAQRVRAAAAYETAYGLTVPPPPVIAENRAELMILITNLLGQNTPA 60

Db 73 LAWLITYAESAHAAQAQVASAAAEFAAFMTVPPAEVAANRALLAALVATNVLGQNTPA 132

Qy 61 IAVNEAYGEMWAGDAAMFGYAAATATATATLLPFEAPEMTSACGLLEQAAVVEASD 120

Db 133 IMATEHYGEMWAGDALMYGYAASSA-AAGRLNLUITPSQTANMAGLAGQAAAVSHAA 191

Qy 121 TAAANQ-----LMNVPOALKQLAQTOCTTPSSKLGGLWKTVP--HRSPISNMVSMA 173

Db 192 ASTVQOVLGSLISNLNPAVMGCFASPLTSAADAAGLGGIIQIEELLGITFVQNAINGAV 251

Qy 174 NHMSENTSGVSWTNTLSMLKGFAPAAAQAVQTAAQNV---RAMSSLGSSISGSSGLGG 230

Db 252 N-----TTAWFWWATIPNAVFLGHAPALNPAVTVAADAAPFAAAAAAGLAHTVTPVG 307

Qy 231 -GVAANLGRAASVRYGHRDGGKYXSGRRNGGPA 263

Db 308 ASLTASLGEASSV-----GGLSVFAGWSTAAPA 335

RESULT 14

B70932

Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C/Accession: B70932

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Fellwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998

A/Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: B70932

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-468 <COL>

A/Cross-references: GB:AL020221; GB:AL123456; NID:93250699; PIDN:CAA17730.1; PID:el254621

A/Experimental source: strain H37RV

C/Genetics:

A/Gene: PPE

RESULT 14
B70932
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C/Species: Mycobacterium tuberculosis
C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C/Accession: B70932
R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Genies, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S. Nature 393, 537-544, 1998
A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230
A/Accession: B70932
A/Status: preliminary; nucleic acid sequence not shown; translation not shown
A/Molecule type: DNA
A/Residues: 1-468 <COL>
A/Cross-references: GB:AL020201; GB:AL123456; NID:93250699; PID:CAA17730.1; PID:el254620
A/Experimental source: strain H37RV
C/Genetics:
A/Gene: PPE

Job time : 7.87361 secs

```

Query Match      29.6%; Score 386.5; DB 2; Length 468;
Best Local Similarity 37.8%; Pred. No. 4.1e-19;
Matches 102; Conservative 37; Mismatches 84; Indels 47; Gaps 7;

QY      1 VAWMSVTAGQAELTAAQVRVAAAAYETAYGLTVPPVIAENRAELMILIATNLLGQNTPA 60
Db      73 VTMWSGTSAQAKAADQARAADVAYETAFRAVVPVPPQIAANRSQLISLVATNIFGNTAA 132

QY      61 IAVNEAEYGEWMAQDAAMGYAAATATATATILLPEEAPEMTSAGLLEQAAAYEEAS- 119
Db      133 IATAEAEYGEWMAQDTWAMFGYASSATA-SRUTPTTAPPQTTPNPSGLAQAAATGQATA 191

QY      120 -----DTAAANQLMNNVPQALKQAOQTGTFPSSKLGJL-----W 155
Db      192 LASGTNAVTTALSSAAQFPFDIIPTELLGLA--TLSTQVTLQMGOLINAIFGPTGATT 249

QY      156 KTVSPHRSPISNMVSMANNHSMNTNSGVSMNTLSSMLKGFAPARAQAQAVTAAQNGVRA 215
Db      250 QNVFTAAANVTKFTSWANDAMSAFLUGTEKFVF-----WQPPPAPEIPK----- 294

QY      216 MSSLGSSSLGSSGLGGVAAANLGRAASVRYG 245
Db      295 -SSIGAGL---GLRSGLSAGLAHAASAGLG 320

```

RESULT 15

PROBABLE PPE protein - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
 C:Accession: A70646
 R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
 R:Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Genies, S.; Hamlin, N.; Holroyd, S.
 R:Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Soares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:9825987; PMID:9634230
 A:Accession: A70646
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Accession: A70646
 A:Molecule type: DNA
 A:Residues: 1-380 <COL>
 A:Cross-references: GB:283867; GB:AL123456; NID:g3261695; PID:CA806278.1; PID:e291015;
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: PPE

Query Match	29.0%;	Score 379;	DB 2;	Length 380;
Best Local Similarity	36.0%;	Pred. No. 1e-18;		
Matches 108;	Conservative 40;	Mismatches 96;	Indels 56;	Gaps 10;
QY	1	VAMSVTGAQBLTAAQVRVAAAYATAGVLTVPPIAENRAELMILIAITNLLGQNTPA	60	
DB	73	ICWLYTTEAKTQQTATCARAALAAFEQAYAMTLPPVVAANRIQLLALIAITFFQNTAA	132	
QY	61	IAVNAEYGEYENWAOQAAAMFCYAAATATATATLTPPEEAPMTSGAGLLEQAAAVVEASD	120	
DB	133	TAATEAQAEYENWAOQAAAMYATATASA-AAALLTPFPSPQTNPAGLTAQAAAVSQATD	191	
QY	121	TAAANQLMNNYPQALKQLAQP-----TQGT-----PSK	150	
DB	192	PLSL--LIETVQALQALTIPTFPEDFELDAIFAGYATGVQTDVESFVAGTIGAESN	249	
QY	151	LGGL-----WKTVPSPHRSPISNMVMANNHMTNSG-----VSMNTLLSML-----	193	
DB	250	LGLLNVGDENPAEYTPGDFGIGELVSTSPGGVSGASGAGAAVSAGNTVLASVRANIG	309	
QY	194	KGFAFAAAAQAVOTAQNGVYRAMSSL--GSLGSSGLG--GVAANLGRAASV--RYGHR	247	
DB	310	QLSVPPSWAAPSTPVSALSPAGLTLTPGTDVAEHGMPGVPGVPVGAASGLVPYRGV	369	

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:42:21 ; Search time 4.28192 Seconds
(without alignments)
3198.204 Million cell updates/sec

Title: US-09-597-796C-26

Perfect score: 1306

Sequence: 1 VAWMSVTAGQALFRAQVRV.....YCHRDGGKYXSGRRNGGPA 263

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1001.5	76.7	396	1 YD61 MYCTU	Q11031 mycobacteri
2	402	30.8	463	1 Y102 MYCTU	O53951 mycobacteri
3	395	30.2	408	1 YS92 MYCTU	Q10813 mycobacteri
4	255.5	19.6	678	1 YF48 MYCTU	Q10778 mycobacteri
5	243.5	18.6	443	1 Y878 MYCTU	Q10540 mycobacteri
6	243	18.6	487	1 Y442 MYCTU	P42611 mycobacteri
7	194.5	14.9	463	1 Y096 MYCTU	Q10892 mycobacteri
8	191.5	14.7	408	1 SRA MYCLE	O07297 mycobacteri
9	158.5	12.1	434	1 YU18 MYCTU	P31500 mycobacteri
10	158	12.1	435	1 YU21 MYCTU	O53268 mycobacteri
11	156	11.9	178	1 YZ29 MYCTU	O06246 mycobacteri
12	152	11.6	176	1 YZ25 MYCTU	Q50703 mycobacteri
13	114.5	8.8	1211	1 BUN2 DROME	Q24523 drosophila
14	112	8.6	2090	1 N214 HUMAN	P35658 homo sapien
15	110.5	8.5	938	1 TRG1 ECOLI	P33790 escherichia
16	107.5	8.2	354	1 YAGU SCHPO	Q10169 schizosacch
17	106	8.1	1783	1 RAA3 CHLRE	P47033 saccharomyc
18	103.5	7.9	881	1 PRY3 YEAST	P76072 escherichia
19	102	7.8	1120	1 STFR ECOLI	P32768 saccharomyc
20	100.5	7.7	1537	1 FLO1 YEAST	Q36110 saccharomyc
21	99	7.6	329	1 PRV2 YEAST	Q04893 saccharomyc
22	99	7.6	659	1 SFRG HUMAN	P40472 saccharomyc
23	99	7.6	1140	1 YN96 YEAST	Q61572 mus musculu
24	98.5	7.5	475	1 SIM1 YEAST	Q10873 mycobacteri
25	98.5	7.5	553	1 FXC1 MOUSE	P33240 homo sapien
26	98.5	7.5	558	1 YJ83 MYCTU	P14959 homo sapien
27	98.5	7.5	577	1 CST2 HUMAN	Q9ul36 homo sapien
28	98.5	7.5	743	1 PO21 HUMAN	P28284 herpes simp
29	98.5	7.5	1075	1 FLO5 YEAST	Q01981 emericella
30	98.5	7.5	1845	1 Z236 HUMAN	Q63850 mus musculu
31	98	7.5	825	1 ICPO HSV2H	
32	97.5	7.5	416	1 CREA EMENI	
33	97.5	7.5	526	1 NU62 MOUSE	

ALIGNMENTS

RESULT 1

YD61 MYCTU STANDARD; PRT; 396 AA.
AC Q11031;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rv1361C/MT1406.
GN Rv1361C OR MT1406 OR MTCY02B10.25C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala E., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares R., Sultston J.E., Taylor K., Whitehead S., Barrrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22208494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
RC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z75555; CAA99966.1; --
DR EMBL; AE007013; AAK45669.1; --
DR FIC; H70741; H70741.
DR TIGR; MT1406; --
DR Tuberculist; Rv1361C; --
DR InterPro; IPR000030; Microbac_PPE.

34 96.5 7.4 461 1 US45 LALCLC
35 96.5 7.4 461 1 Y514 HUMAN
36 96.5 7.4 632 1 PO21 RAT
37 96 7.4 939 1 SLAP CAUFE
38 96 7.4 1025 1 SLAP CAUFE
39 95.5 7.3 2035 1 HFC1 HUMAN
40 95.5 7.3 2090 1 HFC1 MESAU
41 95 7.3 323 1 JUND CHICK
42 95 7.3 774 1 STF LAMED
43 95 7.3 1556 1 PROS DROVI
44 94.5 7.2 601 1 PDM1 DROME
45 94.5 7.2 780 1 CTPA MYCLE

P22865 lactococcus
O60289 homo sapien
P31503 rattus norv
P35827 campylobact
P35828 campylobact
P51610 homo sapien
P51611 mesocricetu
P27921 gallus gall
P03784 bacterioph
Q9661 drosophila
P31368 drosophila
P46839 mycobacteri

```

DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 158 159 TA -> AT (IN REF. 2).
SQ SEQUENCE 396 AA; 40015 MW; 6AFAB0D7B5F668D0 CRC64;

Query Match 76.7%; Score 1001.5; DB 1; Length 396;
Best Local Similarity 84.1%; Pred. No. 2.4e-62;
Matches 207; Conservative 13; Mismatches 21; Indels 5; Gaps 2;

QY 1 VAMSVTAGQALTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLGQNTPA 60
DB 74 VAMSVTAGQALTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLGQNTPA 133
QY 61 IAVNEAEYGENWAQDAAMFGYAATATATATLLPPEAEPMTSAGGLLEQAAVEASD 120
DB 134 IAVNEAEYGENWAQDAAMFGYAATATATATLLPPEAEPMTSAGGLLEQAAVEASD 193
QY 121 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLWKTVSPPHRSPIGNMVMANNHSMTN 180
DB 194 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLWKTVSPPHRSPIGNMVMANNHSMTN 253
QY 181 SGVSMATLSMLKGFAPAAAQAQVTAQNGVRAMSS-----LGSSLGSSGLGGVAAAL 236
DB 254 SGVSMATLSMLKGFAP-AAAQAVETAQNGVRQAAMSSLSGLSSGLGGVAAAL 312
QY 237 GRAASV 242
DB 313 GRAASV 318

RESULT 2
YI02 MYCTU STANDARD; PRT; 463 AA.
AC 053951;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rv1851/MT1851/MB1830.
GN Rv1802 OR MT1851 OR MT049.24 OR MB1830.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Teklaia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwin M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Unayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL J. Bacteriol. 184:5479-5490(2002).
RN [3]
RP SEQUENCE FROM N.A.

```

```

RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrall B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RA "The complete genome sequence of Mycobacterium bovis."
RT Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
RL -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL022021; CAAL17723.1; -
DR EMBL; AE007044; AAK46123.1; -
DR EMBL; BX248340; CAD94533.1; -
DR PIR; C70931; C70931.
DR TIGR; MT1851; -
DR TubercuList; Rv1802; -
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 401 401 S -> L (IN REF. 2).
SQ SEQUENCE 463 AA; 46021 MW; EE64828BF09FA551 CRC64;

Query Match 30.8%; Score 402; DB 1; Length 463;
Best Local Similarity 35.7%; Pred. No. 8.2e-21;
Matches 114; Conservative 37; Mismatches 106; Indels 62; Gaps 8;

QY 1 VAMSVTAGQALTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLGQNTPA 60
DB 72 VAMSVTAGQALTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLGQNTGA 131
QY 61 IAVNEAEYGENWAQDAAMFGYAATATATATLLPPEAEPMTSAGGLLEQAAVEASD 120
DB 132 IAAAEARYAEMWAQDAAMFGYAATATATATLLPPEAEPMTSAGGLLEQAAVEASD 190
QY 121 TAAANQLMNNVPOALKOLAQPTQGTTPSSKLGGLWKTVS--PHRSPI-----SN 167
DB 191 ASAGN-ARSLVSEVLEFLA--TAGTNTNKTIVASLNNAVTGVFVASSVYNSVMLGLGFASK 247
QY 168 MVSMANNHM-----SMTNSGVSMNTLTSSMLKGFAPAA 200
DB 248 MYLPANDTVISTIFGMVQFQKFNPNVTFPNPDLIPKLSALGAGLGRSAISSGLGSTAPAI 307
QY 201 AAQAVQTAQNGVR-----AMSSLSGLSSGLGGVAAANLGRA-----ASV 242
DB 308 SAGASQAGSVGMSVPPSAAATPAIRTAAVFSTGLQAVPAAAISEGSLLSQMALASV 367
QY 243 RYGRHDSKXVXSGRRNGG 261
DB 368 AGGALGGAARATGGLG 386

RESULT 3
YS92 MYCTU STANDARD; PRT; 408 AA.
AC Q10813;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rv2892c/MT2959/MB2916c.
GN Rv2892c OR MT2959 OR MTCY274.23C OR MB2916C.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;

```

[illegible]

FT CONFLICT 258 258 D -> G (IN REF. 2).
SQ SEQUENCE 678 AA; 66736 MW; 209F1593D52533A2 CRC64;
Query Match 19.6%; Score 255.5; DB 1; Length 678;
Best Local Similarity 27.3%; Pred. No. 1.7e-10;
Matches 72; Conservative 34; Mismatches 97; Indels 61; Gaps 6;
CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

CC EMBL; 273101; CAA97385.1; --
DR EMBL; AE006977; RAK45143.1; ALT_INIT.
DR PIR; C70780; C70780.
DR TIGR; MT0901; --
DR Tuberculist; RV0878c; --
DR InterPro; IPR00030; Microbac_PPE.
DR InterPro; IPR002989; Mycobac_Pentapep.
DR Pfam; PF01469; Pentapeptide_2; 4.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Transmembrane; Repeat; Complete proteome.
FT TRANSMEM 15 35 POTENTIAL.
FT TRANSMEM 38 58 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 181 201 POTENTIAL.
FT DOMAIN 64 73 POLY-ALA.
FT DOMAIN 81 115 ALA-RICH.
FT DOMAIN 231 270 4 X 10 AA APPROXIMATE REPEATS.
SQ SEQUENCE 443 AA; 43592 MW; C58BEC607F0675E2 CRC64;
Query Match 18.6%; Score 243.5; DB 1; Length 443;
Best Local Similarity 29.3%; Pred. No. 7.2e-10;
Matches 82; Conservative 38; Mismatches 99; Indels 61; Gaps 12;
CC 1 VAWMSVTAGQAEITAAQVRVAAAYETAYGLTVPPVIAENRAELMLLIATNLGQNTPA 60
DB 76 LSWLNAATARAEGAAAGAKAAAYEAARAATAHPALVAANRQLLSVLNLFQGNLPA 135
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPENTSGAGLLEQAAAVEASD 120
DB 136 IAAEASYEQLWAQDAAMVGVYHGAGSTVASQLTFWQO-----LLSVLPVVTAAAP 186
QY 121 TAAANQLMNVPAQLKQLAQPTQTTPSSKLGGLWKTVPSPHRSPISNV---SNANHMS 177
DB 187 AGAV-----GVFAA---LAIPALGV---ENIG-----VGNFLGIGNIGNNVG 223
QY 178 MTNSG-----VSMNTLTSSMLKGFAPAAAOVQTAQAQGVAMSSLSGS-----SL 223
DB 224 SGTGDNVYFGNIGNAN--LGNIGNANLGSNAGFFNGNDGNTNFSGNAGFLNI 282
QY 224 GSSGLGGVAANLGRAASVRYGHR--DGGKYXSGRNGG 261
DB 283 GSGNEGSG---NLG-----FGNAGDDNTGWSGDTNTG 313
RESULT 6
Y442 MYCTU STANDARD; PRT; 487 AA.
AC P42611; OS3727;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV0442C/MT0458.
OS RV0442C OR MT0458 OR MT037.06C.
OC Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_taxid=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Erdmann;
RC MEDLINE=87137260; PubMed=3029018;
RT "The 65-kilodalton antigen of Mycobacterium tuberculosis."
RL J. Bacteriol. 169:1080-1088 (1987).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=H37Rv;
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

FT CONFLICT 258 258 D -> G (IN REF. 2).
SQ SEQUENCE 678 AA; 66736 MW; 209F1593D52533A2 CRC64;
Query Match 19.6%; Score 255.5; DB 1; Length 678;
Best Local Similarity 27.3%; Pred. No. 1.7e-10;
Matches 72; Conservative 34; Mismatches 97; Indels 61; Gaps 6;
CC the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

CC EMBL; 273101; CAA97385.1; --
DR EMBL; AE006977; RAK45143.1; ALT_INIT.
DR PIR; C70780; C70780.
DR TIGR; MT0901; --
DR Tuberculist; RV0878c; --
DR InterPro; IPR00030; Microbac_PPE.
DR InterPro; IPR002989; Mycobac_Pentapep.
DR Pfam; PF01469; Pentapeptide_2; 4.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Transmembrane; Repeat; Complete proteome.
FT TRANSMEM 15 35 POTENTIAL.
FT TRANSMEM 38 58 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 181 201 POTENTIAL.
FT DOMAIN 64 73 POLY-ALA.
FT DOMAIN 81 115 ALA-RICH.
FT DOMAIN 231 270 4 X 10 AA APPROXIMATE REPEATS.
SQ SEQUENCE 443 AA; 43592 MW; C58BEC607F0675E2 CRC64;
Query Match 18.6%; Score 243.5; DB 1; Length 443;
Best Local Similarity 29.3%; Pred. No. 7.2e-10;
Matches 82; Conservative 38; Mismatches 99; Indels 61; Gaps 12;
CC 1 VAWMSVTAGQAEITAAQVRVAAAYETAYGLTVPPVIAENRAELMLLIATNLGQNTPA 60
DB 76 LSWLNAATARAEGAAAGAKAAAYEAARAATAHPALVAANRQLLSVLNLFQGNLPA 135
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPENTSGAGLLEQAAAVEASD 120
DB 136 IAAEASYEQLWAQDAAMVGVYHGAGSTVASQLTFWQO-----LLSVLPVVTAAAP 186
QY 121 TAAANQLMNVPAQLKQLAQPTQTTPSSKLGGLWKTVPSPHRSPISNV---SNANHMS 177
DB 187 AGAV-----GVFAA---LAIPALGV---ENIG-----VGNFLGIGNIGNNVG 223
QY 178 MTNSG-----VSMNTLTSSMLKGFAPAAAOVQTAQAQGVAMSSLSGS-----SL 223
DB 224 SGTGDNVYFGNIGNAN--LGNIGNANLGSNAGFFNGNDGNTNFSGNAGFLNI 282
QY 224 GSSGLGGVAANLGRAASVRYGHR--DGGKYXSGRNGG 261
DB 283 GSGNEGSG---NLG-----FGNAGDDNTGWSGDTNTG 313
RESULT 6
Y442 MYCTU STANDARD; PRT; 487 AA.
AC P42611; OS3727;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV0442C/MT0458.
OS RV0442C OR MT0458 OR MT037.06C.
OC Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_taxid=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H37Rv;
RC MEDLINE=87137260; PubMed=3029018;
RT "The 65-kilodalton antigen of Mycobacterium tuberculosis."
RL J. Bacteriol. 169:1080-1088 (1987).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=H37Rv;
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne K., Skelton S., Squares S., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Rogers J.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis clinical
RT laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M15467; AAA88235.1; AUT_INIT.
DR EMBL; AL021932; CAAL7399.1; -.
DR EMBL; A3006948; AAK44681.1; -.
DR PIR; C70830; C70830.
DR TIGR; MT0458; -.
DR TubercuList; RV00458; -.
DR InterPro; IPR000030; Microbac_PPE.
DR InterPro; IPR002989; Mycobac_pentapep.
DR Pfam; PF01469; Pentapeptide_2; 5.
DR Pfam; PF00823; PPE; 1.
DR KX Hypothetical protein; Complete proteome.
FT CONFLICT 40 40 E -> X (IN REF. 2).
FT CONFLICT 96 96 I -> T (IN REF. 1).
FT CONFLICT 211 211 G -> GNNNIG (IN REF. 1).
FT SEQUENCE 487 AA; 47247 MW; 97234DB5316C8C7F CRC64;
SQ
Query Match 18.6%; Score 243; DB 1; Length 487;
Best Local Similarity 28.8%; Pred. No. 8.6e-10;
Matches 83; Conservative 30; Mismatches 129; Indels 46; Gaps 8;
QY 1 VAMSVTAGQAEILTAQVRVAAAYETAYGLTVPPPVIAENRAELMILITATLLGNTFA 60
Db :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
76 LAWLSTAAQAEQAAQAAQAAIAFAFAAALTAQPAVAANRGLMQLLAATNFWGNAFA 135
QY 61 IAVNAEYEGMVAQDAADAAAFGAAATATATATLPPPEAPEMTSAG-----GLLEQAAA 114
Db :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
136 LMDVEAAYEQWALDVAVWAGYHFDASAQAQAPQOV--LRNLGIDGKNGQINLFG 193
QY 115 VEEASPTAANQLMNVVPQALKQAQPTQGTTPSSKLG-----GLWKTVS----- 159
Db :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
194 NTSGSGNIGNNI GNNNIGSGNTGTGNTSGNGLGNLGDGNGIGFTGSGNIGFG 253
QY 160 ---PHRSPTSNVYSMANNHSMNTSGVSMNTLSSMLKGFAPAAAQAQVTAQAQNVRAM 216
Db :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
254 ITGDHQMGFGFGSGGN--TGFNGSGNGVGNLFNS-----GSGNIGNIGNS 297
QY 217 SSLGSLGSSGLGGVAAANIGRAASVRYGHRDGG--KYAXSGRRNGGPA 263
Db :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
298 GSLNSGIGTSGT---INAGLGSAGSLNTSPFNAGMQLAALGSAAGSEA 342
```

RESULT 7

```

Y096 MYCTU STANDARD; PRT; 463 AA.
AC Q10892;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV0096/MT0105.
GN RV0096 OR MT0105 OR MTCY251.15.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne K., Skelton S., Squares S., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Rogers J.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z74410; CAA98932.1; -.
DR EMBL; AE006922; NAK44327.1; -.
DR TIGR; H70750; H70750.
DR TubercuList; RV0096; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 88 108 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 216 236 POTENTIAL.
FT TRANSMEM 245 265 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
FT TRANSMEM 419 439 POTENTIAL.
SQ SEQUENCE 463 AA; 46894 MW; 42D9D66A033D0DD8 CRC64;
Query Match 14.9%; Score 194.5; DB 1; Length 463;
Best Local Similarity 36.3%; Pred. No. 1.8e-06;
```

Matches 53; Conservative 21; Mismatches 55; Indels 17; Gaps 3;
 QY 1 VAWKSVTAQAGELTAQVRAAAYETAYGLTVPPVIAENRAELMILIIATNLGQNTTPA 60
 DB 70 LAWLEQTAINSAVTAQAACHVAHAAYCAALAMPPTFAELAAHAIHGVLIATNFFGINTVP 129
 QY 61 IAVNEABYGEWAAQDAAMFCYAAATATATATLLPFEAPEMTSAGLLLEQAAAVEASD 120
 DB 130 IALNEADYVRWQLQADTMAAYQAVADAATVAVSTQAPPIRAPGG-----DAAD 180
 QY 121 T-----AAANQLMNVNPOALKQLAQP 141
 DB 181 TRLDVLSIGQLIRDI---LDFIANP 203

RESULT 8

SRA_MYCLE STANDARD; PRT; 408 AA.
 ID SRA_MYCLE
 AC Q07297;
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Serine-rich antigen (25k) (45 kDa protein).
 GN SRA OR ML0411 OR MUC1383.14.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteridae; Actinobacteriales; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93239328; PubMed=8478104;
 RA Vega-Lopez F., Brooks L.A., Dockrell H.M., de Smet K.A.L.,
 RA Thompson J.K., Hussain R., Stoker N.G.;
 RT "Sequence and immunological characterization of a serine-rich antigen
 from Mycobacterium leprae."
 RL Infect. Immun. 61:2145-2153(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95020554; PubMed=7934845;
 RA Rinke de Wit T.F., Clark-Curtiss J.E., Abebe F., Kolkon A.H.J.,
 RA Jonson A.A.M., Thole J.E.R.;
 RT "A Mycobacterium leprae-specific gene encoding an immunologically
 recognized 45 kDa protein."
 RL Mol. Microbiol. 10:829-838(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eiglmeyer K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Felwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus."
 RL Nature 409:1007-1011(2001).
 CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U00015; AAC43220.1;
 CC EMBL; X68431; CAA48480.1;
 CC EMBL; Z21952; CAA79350.1;
 CC EMBL; Z97179; CAB09398.1;
 CC -----
 CC DR EMBL; U00015; AAC43220.1;
 CC DR EMBL; X68431; CAA48480.1;
 CC DR EMBL; Z21952; CAA79350.1;
 CC DR EMBL; Z97179; CAB09398.1;

DR EMBL; AL583918; CAC29919.1; -;
 DR PIR; C86960; C86960.
 DR PIR; S33522; S33522.
 DR PIR; S39872; S39872.
 DR Lepidoma; ML0411; -;
 DR InterPro; IPR00030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Antigen; Repeat; Complete proteome.
 FT DOMAIN 182 196
 FT REPEAT 209 235
 FT REPEAT 209 214
 FT REPEAT 230 235
 FT CONFLICT 132 132 T -> S (IN REF. 2).
 FT CONFLICT 139 189 S -> L (IN REF. 2).
 FT CONFLICT 191 191 H -> D (IN REF. 2).
 FT CONFLICT 232 292 P -> L (IN REF. 2).
 SQ SEQUENCE 408 AA; 42466 MW; 5C0C2BE0D6E8A9D8 CRC64;
 Query Match 14.7%; Score 191.5; DB 1; Length 408;
 Best Local Similarity 24.5%; Pred. No. 2.6e-06;
 Matches 70; Conservative 49; Mismatches 104; Indels 63; Gaps 8;
 QY 1 VAWKSVTAQAGELTAQVRAAAYETAYGLTVPPVIAENRAELMILIIATNLGQNTTPA 60
 DB 74 VAWLDGNAENAGLITARVLHAYAFEEARAGWVPLLTVLGNIIHTMALKAINWFGQVSTT 133
 QY 61 IAVNEABYGEWAAQDAAMFCYAAATATATATLLPFEAPEMTSAGLLLEQAAAVEASD 120
 DB 134 VAALADYDLMWQNSNTAMTYRDTVLRETCKMENFEPAQLVSR-YCMDESDSVNFSHS 192
 QY 121 TAANQLMNVNPOALKQLAQP-----PTQ----- 143
 DB 193 SSSSDLYESIDNLYDSVAQEEHGSMSQSYNTCGSVAQSELCDSPFPPTPSQSQSND 252
 QY 144 --GTPPSKLGKMTKVSHPSPISNMVMANNMTNSGVSMNTLSSMLKGFAPAAA 201
 DB 253 LSATSLTQQLGL-----DSIISSASALLTNS--ISSSTASSIM----PIVA 295
 QY 202 AAQVQTAQNGVRAMSLGSSLSGLSGGGVAAAN-----LGRAASV 242
 DB 296 SOVTEITLGRSQV-AVEKMIQSISSSTAVSDVAASKVAGVQAVSV 340

RESULT 9

YU18 MYCTU STANDARD; PRT; 434 AA.
 ID YU18 MYCTU
 AC P31500; O53265;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PFB-family protein RV3018C/MT3098/MT3101.
 GN RV3018C OR MT3098/MT3101 OR MTV012.32C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteridae; Actinobacteriales; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Felwell T., Gentsles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogan A., Mclean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.

```

RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=2206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
RN [3]
RP SEQUENCE OF 160-374 FROM N.A.
RX STRAIN=Isolate 50410;
RA Patki A.H., Dale J.W.;
RL Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -!- CAUTION: In strain Oshkosh the gene for this protein is
CC interrupted in position 307 by an IS6110 element.
CC -!- CAUTION: Was originally (Ref.3) thought to be a dihydrofolate
CC reductase.
CC -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC in positions 294; 337 and 355.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL021287; CAAL6103.1; -
DR EMBL; AS007129; AAK47427.1; ALT_SEQ.
DR EMBL; AS007129; AAK47430.1; ALT_SEQ.
DR EMBL; X59271; CAAL1961.1; ALT_FRAME.
DR PIR; E70857; E70857.
DR TIGR; MT3098; -.
DR TIGR; MT3101; -.
DR TubercuList; RV3018c; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 434 AA; 43029 MW; 41D673C4BD389DD6 CRC64;

Query Match 12.1%; Score 158.5; DB 1; Length 434;
Best Local Similarity 23.4%; Pred. No. 0.00053;
Matches 67; Conservative 34; Mismatches 130; Indels 55; Gaps 7;

QY 1 VAMSVTAGQBELTAAQVRVAAAYETAYGLTVPPVIAENRAELMILIAITNLLGQNTPA 60
DB 76 VAWLVQASADSAAAGHEHEAAAGYVCALAEMPTLPDLAANHLTHAVLVATNFFGINTIP 135
QY 61 IAVNEAEYGENWAQDAAMFGAAATATATALLPPEAPEMTSAGGLEQAAVEEASD 120
DB 136 IALNEADYVRMWVQAATVMSAYEAVVGAALVATPHTGPAPVIVKPG-----ANEASN 187
QY 121 TAAAN-----QLMNVFQALKQALQPTQGTTPSKGLGKWTSPH---RSPTS 166
DB 188 AVAAAIITPPWHEIVQFLEETPAAYDQVLSALLSLPA--VAVWFQFLVDILGFNIIG 245
QY 167 NMVSMANNHSMVNTSGVS-----MNTNLSMMLKGPA 197
DB 246 FIITLASNAQLLEFAINASYAVAGLLYATAGVIDIWEVWIGNLFGVYLLGLLGLAL 305
QY 198 PAAQAQVQ--TAQNGVRAMSSIGSSIGS--SGLGGGVAANIGRAAS 241
DB 306 AAAVVPGVAGLVAGVLAUPAVGAAGAPALVGSVAPVSGGVVS 351

RESULT 10
YU21_MYCTU
ID_YU21_MYCTU STANDARD; PRT; 435 AA.
AC OS3268; OS3269;

```

```

DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV3022c/MT3106.
GN RV3022C/MT3106 OR MT3106 OR MTV012.35C/MTV012.36C.
OS Mycobacterium tuberculosis
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H37RV;
RC MEDLINE=9829587; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Ruster S., Seeger K., Skelton S., Squares R.,
RA Stulton J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=CDC 1551 / Oshkosh;
RC MEDLINE=2206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 82.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL021287; CAAL6106.1; ALT_FRAME.
DR EMBL; AL021287; CAAL6107.1; ALT_FRAME.
DR EMBL; AE007129; AAK47435.1; -.
DR TIGR; MT3106; -.
DR TubercuList; RV3021c; -.
DR TubercuList; RV3022c; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 299 299 G -> A (IN REF. 2).
FT CONFLICT 317 320 LAGV -> VTGL (IN REF. 2).
FT CONFLICT 326 326 L -> V (IN REF. 2).
SQ SEQUENCE 435 AA; 42876 MW; 3B157643EAA8484A CRC64;

Query Match 12.1%; Score 158; DB 1; Length 435;
Best Local Similarity 21.9%; Pred. No. 0.00057;
Matches 67; Conservative 32; Mismatches 113; Indels 94; Gaps 7;

QY 1 VAMSVTAGQBELTAAQVRVAAAYETAYGLTVPPVIAENRAELMILIAITNLLGQNTPA 60
DB 76 VAWLVQASADSAAAGHEHEAAAGYVCALAEMPTLPDLAANHLTHAVLVATNFFGINTIP 135
QY 61 IAVNEAEYGENWAQDAAMFGYAA-----ATATATAT 92
DB 136 IALNEADYVRMWVQAATVMSAYEAVVGAALVATPHTGPAPVIVKPGANEASNAVAATIT 195

```

```
OY 93 LLPRFEE-----APENTSAGGLEQAAVEASDTAAANQLMNVPQALKQIAQTQ 144
DB 196 PPFPELAKFLEMAAQAFTEVGEILIMKSAEAMAVGFVELITGLVNFEP----- 243
OY 145 TTPSSKLGGLWKTSRPHSPISNMVSM-----ANNHMSMTNSGV 183
DB 244 -----WLV-----LTMIDMFFATVGFALGVFVLPLEFAVLELAILSIGW 286
OY 184 SMTNTLS-----MLKGFAPAAAAQVQTAAGVGRVMSLSSGLS-SGLGGGVAA 235
DB 287 IISNIFGAIPVLGFLGALAAAVVPGVAGLVAGVLAALFAVGAAGAFAALVGSVAP 346
OY 236 LGRAAS 241
DB 347 SGGVVS 352

RESULT 11
YY29 MYCTU STANDARD; PRT; 178 AA.
AC O06246;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical PPE-family protein RV3429/MT3533.
GN RV3429 OR MT3533 OR MTCY77.01.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gunn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains."
RL J. Bacteriol. 184:5479-5490 (2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z95389; CAB08678.1; -.
CC DR EMBL; AE007158; AAK47873.1; -.
CC DR PIR; C70975; C70975.
CC DR TIGR; MT3533; -.
CC TubercuList; RV3429; -.

OY 93 WMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGONTAIA 62
DB 73 WLSKSSQLKHAANVINGLANAYNDRTRKVVPPPEIAANREERRRLIASNVAGVNTAIA 132
OY 63 VNEAEYGEWMAQDAAMFGYAAATATATATATATATATATATATATATATATATATATAT 106
```

```
DR InterPro: IPR00030; Microbac_PPE.
DR Pfam: PF00823; PPE; 1. Complete proteome.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 178 AA; 19811 MW; 8BE1FC025ABFBEA6 CRC64;

Query Match 11.9%; Score 156; DB 1; Length 178;
Best Local Similarity 37.1%; Pred. No. 0.00029;
Matches 39; Conservative 16; Mismatches 50; Indels 0; Gaps 0;

OY 3 WMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGONTAIA 62
DB 73 WLSKSSQLKHAANVINGLANAYNDRTRKVVPPPEIAANREERRRLIASNVAGVNTAIA 132
OY 63 VNEAEYGEWMAQDAAMFGYAAATATATATATATATATATATATATATATATATATATAT 107
DB 133 GLDAQYQYRAQNTIAVNDYQSTARFILAFLPRWQEPPIYGGGG 177

RESULT 12
YY25 MYCTU STANDARD; PRT; 176 AA.
AC Q50703;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE Hypothetical PPE-family protein RV3425.
GN RV3425 OR MTCY78.04C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544 (1998).
RN -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z77165; CAB01031.1; -.
CC DR PIR; F70738; F70738.
CC TubercuList; RV3425; -.
CC DR InterPro: IPR000030; Microbac_PPE.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 176 AA; 19855 MW; B8CE2E9463B87B0 CRC64;

Query Match 11.6%; Score 152; DB 1; Length 176;
Best Local Similarity 36.5%; Pred. No. 0.00054;
Matches 38; Conservative 17; Mismatches 49; Indels 0; Gaps 0;

OY 3 WMSVTAGQAEILTAQVRVAAAAYETAYGLTVPPVIAENRAELMILLIATNLLGONTAIA 62
DB 73 WLSKSSQLKHAANVINGLANAYNDRTRKVVPPPEIAANREERRRLIASNVAGVNTAIA 132
OY 63 VNEAEYGEWMAQDAAMFGYAAATATATATATATATATATATATATATATATATATATAT 106
```


Db 133 DLDAQDYQYRARNVAVNNAKVSQWTRGALSOLFWRPPOYRG 176

RESULT 13

BUN2_DROME STANDARD; PRT; 1211 AA.

AC Q24523; Q9VK78; Q9VK79;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DE 10-OCT-2003 (Rel. 42, Last annotation update)

DE Bunched protein, class 2/class 3 isoforms (Shortsighted protein).

GN BUN OR SHS OR CG5461.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN SEQUENCE FROM N.A. (ISOFORM CLASS 2), SUBCELLULAR LOCATION, AND FUNCTION

RP FUNCTION

RC TISSUE=Eye-antennal disk;

RX MEDLINE=96038094; PubMed=7555710;

RA Treisman J.E., Lai Z.-C., Rubin G.M.;

RT "Shortsighted acts in the decapentaplegic pathway in Drosophila eye development and has homology to a mouse TGF-beta-responsive gene.";

RL Development 121:2835-2845(1995).

RN [2]

RP SEQUENCE FROM N.A. (ISOFORMS CLASS 2 AND CLASS 3).

RC STRAIN=Berkley;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Vandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champagne M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,

RA April J.F., Agbayani A., An H.-J., Andrews-Franckoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.V., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foslter C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Laoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclet J.M., Pail J.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Furi V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas P., Testor C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong X., Zhou X., Zhu S., Zhu X., Smith H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

CC -!- FUNCTION: Probable transcription factor required for peripheral nervous system morphogenesis, eye development and oogenesis. May be required for the transmission of the app signal and for a morphogenetic movement of the medulla in the brain that reorients

CC the second optic lobe relative to the first. Plays a role in determining proper dorsal cell fates leading to the formation of the dorsal appendages.

CC -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic.

CC -!- ALTERNATIVE PRODUCTS:

CC Event=Alternative splicing; Named isoforms=3;

CC Comment=Experimental confirmation may be lacking for some isoforms;

CC Name=Class 2;

CC IsoId=Q24523-1; Sequence=Displayed;

CC Name=Class 1;

CC IsoId=Q24522-1; Sequence=External;

CC Name=Class 3;

CC IsoId=Q24523-2; Sequence=VSP_006670;

CC -!- SIMILARITY: Belongs to the TSC-22/Dip/Bun family.

CC -!- CAUTION: Ref.2 sequence differs from that shown due to erroneous gene model prediction.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; L42512; AAC41608.1; .

DR EMBL; AE003636; AAF53200.1; ALT_SEQ.

DR EMBL; AE003636; AAF53201.1; ALT_SEQ.

DR PIR; T13804; T13804.

DR HSP; F80220; 1DIP.

DR FlyBase; Fgn00C10460; bun.

DR GO; GO:0007422; P:peripheral nervous system development; NAS.

DR InterPro; IPR000580; TSC-22_Dip_Bun.

DR Pfam; PF01166; TSC22; 1.

DR ProDom; PD007152; TSC-22_Dip_Bun; 1.

DR PROSITE; PS01289; TSC22_1.

KW Transcription regulation; Nuclear protein; Alternative splicing.

FT DOMAIN 15 31 POLY-GLN.

FT DOMAIN 31 86 ALA-RICH.

FT DOMAIN 86 102 POLY-GLN.

FT DOMAIN 102 241 POLY-SER.

FT DOMAIN 241 254 POLY-GLN.

FT DOMAIN 254 265 POLY-SER.

FT DOMAIN 265 321 POLY-HIS.

FT DOMAIN 306 328 POLY-GLN.

FT DOMAIN 328 356 POLY-GLY.

FT DOMAIN 348 619 POLY-GLN.

FT DOMAIN 607 619 POLY-GLN.

FT DOMAIN 661 668 POLY-ALA.

FT DOMAIN 743 746 POLY-ALA.

FT DOMAIN 759 765 POLY-GLN.

FT DOMAIN 795 801 POLY-GLN.

FT DOMAIN 817 822 POLY-GLN.

FT DOMAIN 832 838 POLY-ALA.

FT DOMAIN 884 891 POLY-ALA.

FT DOMAIN 927 947 GLN-RICH.

FT DOMAIN 1001 1005 POLY-ALA.

FT DOMAIN 1011 1014 POLY-ALA.

FT DOMAIN 1069 1090 LEUCINE-ZIPPER.

FT DOMAIN 1194 1201 POLY-ALA.

FT VARSPLIC 1 109 Missing (in isoform Class 3).

FT CONFLICT 10 10 /FTIQ=VSP_006670.

FT CONFLICT 328 328 K -> E (IN REF. 1).

FT CONFLICT 801 801 MISSING (IN REF. 1).

FT CONFLICT 1199 1195 Q -> QQQ (IN REF. 1).

FT CONFLICT 1211 1211 QVTSAA -> TS (IN REF. 2).

SQ SEQUENCE 1211 AA; 125299 MW; 9925A9159A7051B0 CRC64;

Query Match 8.8%; Score 114.5; DB 1; Length 1211;

Best Local Similarity 25.9%; Pred. No. 1.8;

Matches 66; Conservative 28; Mismatches 80; Indels 81; Gaps 13;

QY 14 TAAQVRVAAAAYETAYGLTVPFPVTAENPAELMILLIATNLGQNTFAIA---VNEAEYGE 70

RC STRAIN=K12;
RX MEDLINE=92204127; PubMed=1348105;
RA Firth N., Skurray R.A.;
RT "Characterization of the F plasmid bifunctional conjugation gene,
RT traG.";
RL Mol. Gen. Genet. 232:145-153(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=94359430; PubMed=7915817;
RX Frost L.S., Ippen-Ihler K., Skurray R.A.;
RT "Analysis of the sequence and gene products of the transfer region of
RT the F sex factor.";
RL Microbiol. Rev. 58:162-210(1994).
RN [3]
RP SEQUENCE FROM N.A.
RA Shimizu H., Saitoh Y., Suda Y., Uehara K., Sampei G., Mizobuchi K.;
RT "Complete nucleotide sequence of the F plasmid: its implications for
RT organization and diversification of plasmid genomes.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PLAYS A CRUCIAL ROLE IN DONOR-RECIPIENT CELL
CC INTERACTIONS. REQUIRED FOR TWO STAGES OF THE CONJUGATION PROCESS:
CC FILUS BIOSYNTHESIS AND MATING AGGREGATE STABILIZATION. MAY
CC INTERACT WITH TRAN.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -!- PTM: TRAG* PROBABLY ARISES FROM THE POST-TRANSLATIONAL CLEAVAGE OF
CC TRAG.
CC -!- CAUTION: TRAG IS NOT RESPONSIBLE FOR THE N-TERMINAL ACETYLATION
CC OF F PILIN AS STATED BY SOME AUTHORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M59763; AAC98081.1; -;
DR EMBL; U01159; AAC44184.1; -;
DR EMBL; AF001918; BAA7969.1; -;
DR EIR; S20480; S20480.
DR Ecogene; EG40114; traG.
KW Plasmid; Conjugation; Transmembrane; Inner membrane;
KW Complete proteome.
FT CHAIN 1 938 TRAG PROTEIN.
FT CHAIN ? 938 TRAG* PROTEIN.
FT DOMAIN 1 53 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 54 73 POTENTIAL.
FT DOMAIN 74 329 PERIPLASMIC (POTENTIAL).
FT TRANSMEM 330 350 POTENTIAL.
FT DOMAIN 351 412 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 413 433 POTENTIAL.
FT DOMAIN 434 938 PERIPLASMIC (POTENTIAL).
SQ SEQUENCE 938 AA; 102471 MW; 9E09C8402ACDFDF CRC64;

Query Match 8.5%; Score 110.5; DB 1; Length 938;
Best Local Similarity 22.8%; Pred. No. 2.6;
Matches 67; Conservative 40; Mismatches 122; Indels 65; Gaps 14;

QY 11 AELTRAQVAAVAAAYETAYGLTVPPV-----IAENRAELMILITNLLGQNTPA--- 60
DB 395 SELSQLKLYSLNASTAGLYSLMIPPLSWVMYKGLGAGFSSVYSHFASSTISPTASAAGS 454
QY 61 IAVNEAEYGE-----WAQDAAMFGY-----AAATATATATLLPFEAPEMTSA 105
DB 455 VVDGNYSGNMQTEENVNGFSWSTNSTTSPQMMYQTGSGATATQTR-----DGNMVMDS 509
QY 106 GGLLEQAAAVEASDTAAANOLM-----NNVQALKQLAQPTGTTTPSKLGLWKTYS- 159
DB 510 GAMSRLPVGINATRTQAAQQENAREASNRBSALHGF-----SSSIASAWNTLSQ 560
QY 160 --PHRSPISNMVSMNNHSMNTNSGVSMNTNTLSSMLKGFAPA---AAQAVQTAAQNGVR 214

Db 561 FGSNRGSSDVTGGADSTMSAQDS--WMASTRVESAVESYAKAHNISNEQATRELASRSTN 618
QY 215 AMSSLG-----SSLGSS--GLGGVAAVLGRAASVRYGHRDGGKYAXSGRR 258
Db 619 A--SLGLYGDVAKGHLGISVLGNGGGVGLQAGAKASIDGSDLDLSHE--ASSGSR 669

Search completed: June 30, 2004, 16:49:48
Job time : 6.28192 secs

the first of these is the fact that the system is not a simple one, but a complex one, in which the various parts are interrelated and interdependent. The second is that the system is not a static one, but a dynamic one, in which the various parts are constantly changing and evolving. The third is that the system is not a closed one, but an open one, in which the various parts are constantly interacting with the environment.

The first of these is the fact that the system is not a simple one, but a complex one, in which the various parts are interrelated and interdependent.

The second is that the system is not a static one, but a dynamic one, in which the various parts are constantly changing and evolving.

The third is that the system is not a closed one, but an open one, in which the various parts are constantly interacting with the environment.

The first of these is the fact that the system is not a simple one, but a complex one, in which the various parts are interrelated and interdependent.

The second is that the system is not a static one, but a dynamic one, in which the various parts are constantly changing and evolving.

The third is that the system is not a closed one, but an open one, in which the various parts are constantly interacting with the environment.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:01 ; Search time 19.156 Seconds
(without alignments)
4331.879 Million cell updates/sec

Title: US-09-597-796C-26

Perfect score: 1306
Sequence: 1 VAMSVTAGQAEITAAQVRV.....YGHROGKXVXSGRRNGGPA 263

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_25: *
1: sp_archaea: *
2: sp_bacteria: *
3: sp_fungi: *
4: sp_human: *
5: sp_invertebrate: *
6: sp_mammal: *
7: sp_mhc: *
8: sp_organelle: *
9: sp_phase: *
10: sp_plant: *
11: sp_rodent: *
12: sp_virus: *
13: sp_vertebrate: *
14: sp_unclassified: *
15: sp_rvirus: *
16: sp_bacteriap: *
17: sp_archaeap: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	DB ID	Description
1	1187	90.9	391	16	O05298	O05298 mycobacteri
2	1158.5	88.7	390	16	Q7U0E9	Q7U0E9 mycobacteri
3	1122.5	85.9	396	16	Q7U071	Q7U071 mycobacteri
4	1008.5	77.2	393	16	Q8VIZ3	Q8VIZ3 mycobacteri
5	949.5	72.7	393	16	O06341	O06341 mycobacteri
6	949.5	72.7	393	16	Q7TW55	Q7TW55 mycobacteri
7	463	35.5	399	16	Q7TZW8	Q7TZW8 mycobacteri
8	460	35.2	403	16	O53956	O53956 mycobacteri
9	441	33.8	423	16	O53950	O53950 mycobacteri
10	441	33.8	423	16	Q7TIZ4	Q7TIZ4 mycobacteri
11	432.5	33.1	393	16	Q7TZJ3	Q7TZJ3 mycobacteri
12	432.5	33.1	411	16	O53939	O53939 mycobacteri
13	424.5	32.5	421	16	Q9Z5K0	Q9Z5K0 mycobacteri
14	422	32.3	409	16	O53957	O53957 mycobacteri
15	420	32.2	409	16	Q7TZW7	Q7TZW7 mycobacteri
16	420	32.2	410	2	Q99Q11	Q99Q11 mycobacteri

17	403	30.9	413	16	O06386	O06386 mycobacteri
18	402	30.8	391	16	P96362	P96362 mycobacteri
19	402	30.8	391	16	Q7U075	Q7U075 mycobacteri
20	390	29.9	394	16	Q7TXX5	Q7TXX5 mycobacteri
21	389	29.8	462	16	O33310	O33310 mycobacteri
22	387.5	29.7	694	16	Q8VJW0	Q8VJW0 mycobacteri
23	386.5	29.6	468	16	O53958	O53958 mycobacteri
24	379	28.0	380	16	P95190	P95190 mycobacteri
25	379	28.0	380	16	Q7TX66	Q7TX66 mycobacteri
26	375	28.7	385	16	O33204	O33204 mycobacteri
27	375	28.7	385	16	Q8VJZ0	Q8VJZ0 mycobacteri
28	375	28.7	385	16	Q7TZR7	Q7TZR7 mycobacteri
29	361	27.6	381	16	Q7TX67	Q7TX67 mycobacteri
30	355	27.2	382	16	Q7TXK3	Q7TXK3 mycobacteri
31	355	27.2	397	2	Q9AGF0	Q9AGF0 mycobacteri
32	352	27.0	402	16	O33312	O33312 mycobacteri
33	348.5	26.7	365	16	O86373	O86373 mycobacteri
34	347.5	26.6	405	16	Q8VJW5	Q8VJW5 mycobacteri
35	347	26.6	364	16	Q7TZJ5	Q7TZJ5 mycobacteri
36	346.5	26.5	394	16	O33205	O33205 mycobacteri
37	346.5	26.5	394	16	Q7TZR6	Q7TZR6 mycobacteri
38	346	26.5	350	16	Q7TZJ2	Q7TZJ2 mycobacteri
39	346	26.5	363	16	O53940	O53940 mycobacteri
40	339.5	26.0	443	16	Q8VKU9	Q8VKU9 mycobacteri
41	333.5	25.5	443	16	Q7U242	Q7U242 mycobacteri
42	332.5	25.5	187	16	Q7TZH6	Q7TZH6 mycobacteri
43	329	25.2	391	16	O05798	O05798 mycobacteri
44	328	25.2	391	16	Q7TX76	Q7TX76 mycobacteri
45	328	25.1	406	16	P71869	P71869 mycobacteri

ALIGNMENTS

RESULT 1

O05298 PRELIMINARY; PRT; 391 AA.
ID O05298
AC O05298;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (PPE family protein).
GN RV1196 OR MTC1364.08 OR MT1234.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers R., Rutter S., Seeger K., Skelton S., Squares R., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the Biology of Mycobacterium tuberculosis from the complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O., Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E., Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L., Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A., Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains.";
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

DE PPE family protein.
GN MT3582.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; AE007161; AAK47941.1; -.
DR TIGR; MT3582; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000030; Microbac_PPE.
DR InterPro; IPR000508; Peptidase_S26.
DR Pfam; PF00823; PPE; 1.
DR PROSITE; PS00501; SPASE_1; 1.
SQ SEQUENCE 393 AA; 39688 MW; 86F0B67798855511 CRC64;
Query Match 77.2%; Score 1008.5; DB 16; Length 393;
Best Local Similarity 84.6%; Pred. No. 6.6e-61;
Matches 208; Conservative 14; Mismatches 19; Indels 5; Gaps 2;
QY 1 VAMSVTAGQAELTAQVRAAAAYETAYGLTVPPVIAENRAELMILIAIATNLGQNTPA 60
Db 74 VAMSVTAGQAELTAQVRAAAAYETAYGLTVPPVIAENRAELMILIAIATNLGQNTPA 133
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGLLEQAAVEEASD 120
Db 134 IAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGLLEQAAVEEASD 193
QY 121 TAANQLMNNVPQALQKLAQPTGTTSSKLGGLWKTVPSPISNMVSMANNHSMWN 180
Db 194 TAANQLMNNVPQALQKLAQPTGTTSSKLGGLWKTVPSPISNMVSMANNHSMWN 253
QY 181 SGVSMNTLSSMLKGFAPAAAQAVTAQNGVRAMSS---LGSSLSGSGLGAGVAANL 236
Db 254 SGVSMNTLSSMLKGFAPAAAQAVTAQNGVRAMSS---LGSSLSGSGLGAGVAANL 312
QY 237 GRAASV 242
Db 313 GRAASV 318
RESULT 5
ID O06341 PRELIMINARY; PRT; 393 AA.
AC O06341;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Hypothetical protein RV3478.
GN RV3478 OR MTC13B12.31.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,

RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL; Z95390; CAB08702.1; -.
DR FIR; C70568; C70568.
DR Tuberculist; RV3478; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;
Query Match 72.7%; Score 949.5; DB 16; Length 393;
Best Local Similarity 79.7%; Pred. No. 7e-57;
Matches 196; Conservative 14; Mismatches 31; Indels 5; Gaps 2;
QY 1 VAMSVTAGQAELTAQVRAAAAYETAYGLTVPPVIAENRAELMILIAIATNLGQNTPA 60
Db 74 VAMSVTAGQAELTAQVRAAAAYETAYGLTVPPVIAENRAELMILIAIATNLGQNTPA 133
QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGLLEQAAVEEASD 120
Db 134 IAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPEMTSAGLLEQAAVEEASD 193
QY 121 TAANQLMNNVPQALQKLAQPTGTTSSKLGGLWKTVPSPISNMVSMANNHSMWN 180
Db 194 TAANQLMNNVPQALQKLAQPTGTTSSKLGGLWKTVPSPISNMVSMANNHSMWN 253
QY 181 SGVSMNTLSSMLKGFAPAAAQAVTAQNGVRAMSS---LGSSLSGSGLGAGVAANL 236
Db 254 SGVSMNTLSSMLKGFAPAAAQAVTAQNGVRAMSS---LGSSLSGSGLGAGVAANL 312
QY 237 GRAASV 242
Db 313 GRAASV 318
RESULT 6
ID Q7TWFS PRELIMINARY; PRT; 393 AA.
AC Q7TWFS;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE PE family protein.
GN PPE60 OR MB3505.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12789972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248346; CAD95692.1; -.
KW Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;
Query Match 72.7%; Score 949.5; DB 16; Length 393;
Best Local Similarity 79.7%; Pred. No. 7e-57;
Matches 196; Conservative 14; Mismatches 31; Indels 5; Gaps 2;
QY 1 VAMSVTAGQAELTAQVRAAAAYETAYGLTVPPVIAENRAELMILIAIATNLGQNTPA 60

Db 284 KIGGKPTGALAPLAEFALHTPIILGSEGLGGSVSAGIGRAGLV 327

RESULT 9

OS3950
ID OS3950 PRELIMINARY; PRT; 423 AA.
AC OS3950;
DT 01-JUN-1998 (TREMELrel. 06, Created)
DT 01-JUN-1998 (TREMELrel. 06, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE PPE family protein.
DE RV1801 OR MT1850.
GN MYCobacterium tuberculosis.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RC MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Unayam L.A., Emdorova M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL020201; CAAL7722.1; -.
DR EMBL; AE007044; AAK46122.1; -.
DR PIR; B70931; B70931.
DR TIGR; MT1850; -.
DR TubercuList; Rv1801; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Complete proteome.
SQ SEQUENCE 423 AA; 41477 MW; 26E52CC271FBBF57 CRC64;

Query Match 33.8%; Score 441; DB 16; Length 423;
Best Local Similarity 41.2%; Pred. No. 3.5e-22;
Matches 117; Conservative 30; Mismatches 67; Indels 70; Gaps 9;
QY 1 VAWMSVTAGQELTAQVRAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
Db 72 VAWMSATVARELAGAQAARLAIAYEAFAATVPPVIAANRAQLMVLIAINIFGQNTPA 131
QY 61 IAVNEAFYEGEMWAQDRAAMFYGAATATATATATLPPFEAPMTSAGGLLEQAAVEASD 120
Db 132 IMVTEAQYMEWMAQDRAAMFYGAATATATATLPPFEAPMTSAGGLLEQAAVEASD 190
QY 121 TAAAN-----OLMNVNVPQALKOLAQPT-----QGTFP-----SSKLGGLWKTVP 160
Db 191 TAAGGNLOSAFPQLLSAVPRALQGLAFTASQSASATPQWVTDLGNLSTFLGG--AVTGP 248
QY 161 HRSPISNNVSMANNHMTNSGVSMTNTLSSMLKGFAPAAAAQAVTAQAQNGVRAMSSLG 220
Db 249 YTFP-----GVLPSPGVYLLIGISVL-----VTQNGQGVSAALLG 283
QY 221 S-----SIGSSGLGGG-VANLGRAASV 242

RESULT 11

QY72J3
ID QY72J3 PRELIMINARY; PRT; 393 AA.
AC QY72J3;
DT 01-OCT-2003 (TREMELrel. 25, Created)
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE PPE family protein.
DE PPE26 OR MB1817.
GN Mycobacterium bovis.
OS Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,

Db 284 KIGGKPTGALAPLAEFALHTPIILGSEGLGGSVSAGIGRAGLV 327
RESULT 10
QY72I4
ID QY72I4 PRELIMINARY; PRT; 423 AA.
AC QY72I4;
DT 01-OCT-2003 (TREMELrel. 25, Created)
DT 01-OCT-2003 (TREMELrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMELrel. 25, Last annotation update)
DE PPE family protein.
GN PPE29 OR MB1829.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL; BX248340; CAD94532.1; -.
KW Complete proteome.
SQ SEQUENCE 423 AA; 41477 MW; 26E52CC271FBBF57 CRC64;

Query Match 33.8%; Score 441; DB 16; Length 423;
Best Local Similarity 41.2%; Pred. No. 3.5e-22;
Matches 117; Conservative 30; Mismatches 67; Indels 70; Gaps 9;
QY 1 VAWMSVTAGQELTAQVRAAAYETAYGLTVPPVIAENRAELMILIAINLLGQNTPA 60
Db 72 VAWMSATVARELAGAQAARLAIAYEAFAATVPPVIAANRAQLMVLIAINIFGQNTPA 131
QY 61 IAVNEAFYEGEMWAQDRAAMFYGAATATATATLPPFEAPMTSAGGLLEQAAVEASD 120
Db 132 IMVTEAQYMEWMAQDRAAMFYGAATATATLPPFEAPMTSAGGLLEQAAVEASD 190
QY 121 TAAAN-----OLMNVNVPQALKOLAQPT-----QGTFP-----SSKLGGLWKTVP 160
Db 191 TAAGGNLOSAFPQLLSAVPRALQGLAFTASQSASATPQWVTDLGNLSTFLGG--AVTGP 248
QY 161 HRSPISNNVSMANNHMTNSGVSMTNTLSSMLKGFAPAAAAQAVTAQAQNGVRAMSSLG 220
Db 249 YTFP-----GVLPSPGVYLLIGISVL-----VTQNGQGVSAALLG 283
QY 221 S-----SIGSSGLGGG-VANLGRAASV 242

RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RT "The complete genome sequence of Mycobacterium bovis";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 DR EMBL; BX248340; CAD94520.1; -.
 KW Complete proteome.
 SQ SEQUENCE 393 AA; 38572 MW; 41944B6E547A8AB0 CRC64;

Query Match 33.1%; Score 432.5; DB 16; Length 393;
 Best Local Similarity 41.0%; Pred. No. 1.2e-21;
 Matches 107; Conservative 38; Mismatches 85; Indels 31; Gaps 7;

QY 1 VAWMSVTAQAGELTAQVRAAAAYETAYGLTVPPVIAENRAELMILIAINLLQNTTPA 60
 DB 73 VAWMSAAAQAQEAQATQARAAAAAFAAFAATVPPPLIAANRASLMQIISTNVFGQNTSA 132
 QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGLLEQAAVEEASD 120
 DB 133 IAAEAQYGEWMAQDAAMFYAGSSASASA-VTFSTPPOIANPTAAGTQAAAATAAG 191
 QY 121 TAAA--NQLMNVPOALKOLAQTOGTTPSSKLGGLMK-----TVSPH 161
 DB 192 TAQSTLTEMITGLPNALQSLTSPLLQSS-NGPLSLWQLFGTFNFTSISALLTDLPY 250
 QY 162 RPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQNGVRAMSSLS 221
 DB 251 ASFFYNTGLPYFSGMGNFIQAATL-GLIGSAAPAAVA-----AAGDAAGLPLGG 304
 QY 222 SLGSSGLGGVAANLGRAASV 242
 DB 305 MLG-----GGPVAAGLGNRAASV 321

RESULT 12
 OS3939 PRELIMINARY; PRT; 411 AA.
 AC OS3939;
 DT 01-JUN-1998 (TREMBLrel. 06, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE PPE-family protein.
 GN RV1789 OR MT1838 OR MTV049.11.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares R.,
 RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence";
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains";
 RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL022021; CAAL7711.1; ALT_INIT.
 DR EMBL; AE007043; AAK46108.1; -.
 DR PIR; G70929; G70929.
 DR TIGR; MT1838; -.
 DR Tuberculist; Rv1789; -.
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Complete proteome.
 SQ SEQUENCE 411 AA; 40557 MW; 03C90B5E0590B7DA CRC64;

Query Match 33.1%; Score 432.5; DB 16; Length 411;
 Best Local Similarity 41.0%; Pred. No. 1.3e-21;
 Matches 107; Conservative 38; Mismatches 85; Indels 31; Gaps 7;

QY 1 VAWMSVTAQAGELTAQVRAAAAYETAYGLTVPPVIAENRAELMILIAINLLQNTTPA 60
 DB 91 VAWMSAAAQAQEAQATQARAAAAAFAAFAATVPPPLIAANRASLMQIISTNVFGQNTSA 150
 QY 61 IAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGLLEQAAVEEASD 120
 DB 151 IAAEAQYGEWMAQDAAMFYAGSSASASA-VTFSTPPOIANPTAAGTQAAAATAAG 209
 QY 121 TAAA--NQLMNVPOALKOLAQTOGTTPSSKLGGLMK-----TVSPH 161
 DB 210 TAQSTLTEMITGLPNALQSLTSPLLQSS-NGPLSLWQLFGTFNFTSISALLTDLPY 268
 QY 162 RPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQNGVRAMSSLS 221
 DB 269 ASFFYNTGLPYFSGMGNFIQAATL-GLIGSAAPAAVA-----AAGDAAGLPLGG 322
 QY 222 SLGSSGLGGVAANLGRAASV 242
 DB 323 MLG-----GGPVAAGLGNRAASV 339

RESULT 13
 OS25K0 PRELIMINARY; PRT; 421 AA.
 AC OS25K0;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
 DE Putative PPE protein (PPE-family protein).
 GN ML1182 OR MLCB1701.08C.
 OS Mycobacterium leprae.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TN;
 RX MEDLINE=21128732; PubMed=11234002;
 RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
 RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
 RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
 RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
 RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
 RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
 RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
 RA Barrell B.G.;
 RT "Massive gene decay in the leprosy bacillus";
 RL Nature 409:1007-1011(2001).
 DR EMBL; AL049191; CAB39148.1; -.
 DR EMBL; AL583921; CAC31563.1; -.
 DR PIR; H87056; H87056.
 DR Leproma; ML1182; -.
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Complete proteome.
 SQ SEQUENCE 421 AA; 43119 MW; 680FD5FDCB65B4A7 CRC64;

Query Match 32.5%; Score 424.5; DB 16; Length 421;

Wed Jul 7 12:10:35 2004

us-09-597-796c-26.rspt

Page 8

Db : |||| : || |
304 TPLGGGATGGIARAIYV 320

Search completed: June 30, 2004, 16:52:53
Job time : 20.156 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:44:57 ; Search time 22.1761 seconds
(without alignments)
1697.113 Million cell updates/sec

Title: US-09-597-796C-12

Perfect score: 3686

Sequence: 1 MHHHHHTAASDNFQSQGG.....SGGPVWNLGQVGMNTAAS 729

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3686	100.0	729	4	US-09-223-040-2
2	3686	100.0	729	4	US-09-287-849-2
3	2941	79.8	596	4	US-09-287-849-26
4	1967.5	53.4	600	4	US-09-287-849-22
5	1944	52.7	391	3	US-08-818-112-107
6	1944	52.7	391	4	US-08-818-111-102
7	1944	52.7	391	4	US-09-056-556-107
8	1944	52.7	391	4	US-09-072-596-102
9	1944	52.7	391	4	US-09-072-967-107
10	1652.5	44.8	396	3	US-08-818-112-111
11	1652.5	44.8	396	4	US-08-818-111-106
12	1652.5	44.8	396	4	US-09-056-556-111
13	1652.5	44.8	396	4	US-09-072-596-106
14	1652.5	44.8	396	4	US-09-072-967-111
15	1486.5	40.3	359	3	US-08-818-112-109
16	1486.5	40.3	359	4	US-08-818-111-104
17	1486.5	40.3	359	4	US-09-056-556-109
18	1486.5	40.3	359	4	US-09-072-596-104
19	1486.5	40.3	359	4	US-09-072-967-109
20	1182	32.1	358	4	US-09-287-849-8
21	1182	32.1	263	3	US-08-818-112-91
22	1182	32.1	263	4	US-08-818-111-92
23	1182	32.1	263	4	US-09-056-556-91
24	1182	32.1	263	4	US-09-072-596-92
25	1182	32.1	263	4	US-09-072-967-91
26	983	26.7	355	3	US-08-818-112-79
27	983	26.7	355	4	US-08-818-111-80

28 983 26.7 355 4 US-09-056-556-79 Sequence 79, Appl
29 983 26.7 355 4 US-09-072-596-80 Sequence 80, Appl
30 983 26.7 355 4 US-09-072-967-79 Sequence 79, Appl
31 766.5 20.8 400 4 US-08-073-009-126 Sequence 126, App
32 766.5 20.8 400 4 US-09-073-010-126 Sequence 126, App
33 743.5 20.2 231 4 US-09-287-849-28 Sequence 28, Appl
34 743.5 20.2 654 4 US-09-620-412C-341 Sequence 341, App
35 743.5 20.2 654 4 US-09-598-419-341 Sequence 341, App
36 741 20.1 518 4 US-09-620-412C-333 Sequence 333, App
37 741 20.1 518 4 US-09-598-419-333 Sequence 333, App
38 734 19.9 715 4 US-09-620-412C-321 Sequence 321, App
39 734 19.9 715 4 US-09-598-419-321 Sequence 321, App
40 733.5 19.9 715 4 US-09-620-412C-329 Sequence 329, App
41 733.5 19.9 715 4 US-09-598-419-329 Sequence 329, App
42 727 19.7 631 4 US-09-620-412C-325 Sequence 325, App
43 727 19.7 631 4 US-09-598-419-325 Sequence 325, App
44 723.5 19.6 525 4 US-09-556-877-196 Sequence 196, App
45 723.5 19.6 525 4 US-09-620-412C-196 Sequence 196, App

ALIGNMENTS

RESULT 1

US-09-223-040-2
; Sequence 2, Application US/09223040 ✓
; Patent No. 6544522

GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir

; APPLICANT: Alderson, Mark

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses

; FILE REFERENCE: 014058-009010US

; CURRENT APPLICATION NUMBER: US/09/223,040

; CURRENT FILING DATE: 1998-12-30

; NUMBER OF SEQ ID NOS: 10

; SOFTWARE: Patent In Ver. 2.1

; SEQ ID NO 2

; LENGTH: 729

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion

US-09-223-040-2

Query Match 100.0%; Score 3686; DB 4; Length 729;
Best Local Similarity 100.0%; Pred. No. 3.6e-261;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQSQGGGFAIPICQAWAIAQIRSGGSPVHIGPTAFILGLGVVD 60

Db 1 MHHHHHTAASDNFQSQGGGFAIPICQAWAIAQIRSGGSPVHIGPTAFILGLGVVD 60

QY 61 NNGNGARVORVVGSAASLGISTGVDITAVDGPINSATAMADALNGHHHPGDVISTWQ 120

Db 61 NNGNGARVORVVGSAASLGISTGVDITAVDGPINSATAMADALNGHHHPGDVISTWQ 120

QY 121 TKSOGTRTGNVTLAEGPPAEFMDFGALPPRINSARMYAGPGSASLVAAQWMSVASDL 180

Db 121 TKSOGTRTGNVTLAEGPPAEFMDFGALPPRINSARMYAGPGSASLVAAQWMSVASDL 180

QY 181 FSAASAFOSVWGLTVGSGSSAGLMVAASPVVAMSVTAAQAEITAAQVRAAAAYE 240

Db 181 FSAASAFOSVWGLTVGSGSSAGLMVAASPVVAMSVTAAQAEITAAQVRAAAAYE 240

QY 241 TAYGLTVPFPPIAENRAELMILITNLGQNTPAIVNEAFYGEWQAQDAAMFGYAAAT 300

Db 241 TAYGLTVPFPPIAENRAELMILITNLGQNTPAIVNEAFYGEWQAQDAAMFGYAAAT 300

QY 301 ATATATLLPFEEAEPMTSAGCLLECAAAVEASTAAANQLMNVVQALQLOAQPTQTT 360

Db 301 ATATATLLPFEEAEPMTSAGCLLECAAAVEASTAAANQLMNVVQALQLOAQPTQTT 360

Db 301 ATATATLLPFEAPBMTSAGGLLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTGTT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMTNSGVSMTNTLSSMLKGFAPAAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMTNSGVSMTNTLSSMLKGFAPAAAARQAVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSSGGLGGVAAANLGRASVGSLSVPOQAWAANAQVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSLGSSGGLGGVAAANLGRASVGSLSVPOQAWAANAQVTPAARALP 480
QY 481 LTSLSAAERPGQMLGGLPVGQMGARAGGGLSGVLVPPRPYVWPHSPAAAGDIAPPALS 540
Db 481 LTSLSAAERPGQMLGGLPVGQMGARAGGGLSGVLVPPRPYVWPHSPAAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVLTNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVLTNNHVA 600
QY 601 GATDINAFSVGSGGTGYVDVVGDRDQVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGN 660
Db 601 GATDINAFSVGSGGTGYVDVVGDRDQVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDDAIQPGDSGGPVVNGLGQ 720
Db 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDDAIQPGDSGGPVVNGLGQ 720
QY 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

RESULT 2

US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-2
Query Match 100.0%; Score 3686; DB 4; Length 729;
Best Local Similarity 100.0%; Pred. No. 3.6e-261;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHTAASDNFQSQGGQFAIPQAWAIAQIRSGGSPVTHIGTAFILGLGVVD 60
|||||

Db 1 MHHHHHTAASDNFQSQGGQFAIPQAWAIAQIRSGGSPVTHIGTAFILGLGVVD 60
QY 61 NNGNGARVORVVGSAAPASLSLGISTGDVITAVDGAIPNSATAMADALNGHHFGDVISVTWQ 120
Db 61 NNGNGARVORVVGSAAPASLSLGISTGDVITAVDGAIPNSATAMADALNGHHFGDVISVTWQ 120
QY 121 TKSQGTTRGNTVTLAEGPPAEFVWDFGALPPPEINSGARMYAGPGSASLVAAAQMWDSVASDL 180
Db 121 TKSQGTTRGNTVTLAEGPPAEFVWDFGALPPPEINSGARMYAGPGSASLVAAAQMWDSVASDL 180
QY 181 FSAASAFOSVWVGLTVGWSWIGSSAGLWVAASPYVAMSVTAGOAEITAAQVRVAAAAYE 240
Db 181 FSAASAFOSVWVGLTVGWSWIGSSAGLWVAASPYVAMSVTAGOAEITAAQVRVAAAAYE 240
QY 241 TAYGLTVPPPVIENRAELMILIAITNLLGONTPIAVNEAEYGENWAOAAAAMEFYAAAT 300
Db 241 TAYGLTVPPPVIENRAELMILIAITNLLGONTPIAVNEAEYGENWAOAAAAMEFYAAAT 300
QY 301 ATATATLLPFEAPBMTSAGGLLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTGTT 360
Db 301 ATATATLLPFEAPBMTSAGGLLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTGTT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMTNSGVSMTNTLSSMLKGFAPAAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMTNSGVSMTNTLSSMLKGFAPAAAARQAVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSSGGLGGVAAANLGRASVGSLSVPOQAWAANAQVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSLGSSGGLGGVAAANLGRASVGSLSVPOQAWAANAQVTPAARALP 480
QY 481 LTSLSAAERPGQMLGGLPVGQMGARAGGGLSGVLVPPRPYVWPHSPAAAGDIAPPALS 540
Db 481 LTSLSAAERPGQMLGGLPVGQMGARAGGGLSGVLVPPRPYVWPHSPAAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVLTNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVLTNNHVA 600
QY 601 GATDINAFSVGSGGTGYVDVVGDRDQVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGN 660
Db 601 GATDINAFSVGSGGTGYVDVVGDRDQVAVLQLRGAGGLPSAAIGGVAVGEPVWAMGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDDAIQPGDSGGPVVNGLGQ 720
Db 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDDAIQPGDSGGPVVNGLGQ 720
QY 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

RESULT 3

US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1998-02-18

```

; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; US-09-287-849-26

Query Match          79.8%; Score 2941; DB 4; Length 596;
Best Local Similarity 99.8%; Pred. No. 7.6e-207;
Matches 597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 142 MYDFGALPPEINSRMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTGVSMIG 201
Db 9 MYDFGALPPEINSRMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTGVSMIG 68

QY 202 SSAGLMWAAAAPYVAMKSVTTAGQAELTAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 261
Db 69 SSAGLMWAAAAPYVAMKSVTTAGQAELTAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 128

QY 262 LIATNLLIGONTPTAVNPEAEYGENWAGDAANFGYAAAATATATATLTPPEEAPENTSAGG 321
Db 129 LIATNLLIGONTPTAVNPEAEYGENWAGDAANFGYAAAATATATATLTPPEEAPENTSAGG 188

QY 322 LLEQAAAEVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGGLMKTVPFHRSPISN 381
Db 189 LLEQAAAEVEEASDTAAANQLMNNVPQALQQLAQTQGTTPSSKLGGLMKTVPFHRSPISN 248

QY 382 MYSMANNHSMNTSGVSMNTLSSNLKGFPAPAAQCAVQTAQNGVRAMSSLGSSLGSSG 441
Db 249 MYSMANNHSMNTSGVSMNTLSSNLKGFPAPAAQCAVQTAQNGVRAMSSLGSSLGSSG 308

QY 442 LGGGVAANLGRAASVGSLSVPQAAWAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 501
Db 309 LGGGVAANLGRAASVGSLSVPQAAWAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 368

QY 502 GQMGARAGGGLSGVLRVPPRPYVMPHSPFAGDITAPPALSDREADFPALPLDPSAMVAQV 561
Db 369 GQMGARAGGGLSGVLRVPPRPYVMPHSPFAGDITAPPALSDREADFPALPLDPSAMVAQV 428

QY 562 GPQVNNINTKLYGNNNAVAGAGTGIVIDPNGVVLTNNHVIAGATDINAFSVGSGQTYGVDDV 621
Db 429 GPQVNNINTKLYGNNNAVAGAGTGIVIDPNGVVLTNNHVIAGATDINAFSVGSGQTYGVDDV 488

QY 622 GYDRTQDVAVLQLRGAGGLPSAAIGGGVANGCEPVVAMGNSGGGGTTPRAVGRVVALGQT 681
Db 489 GYDRTQDVAVLQLRGAGGLPSAAIGGGVANGCEPVVAMGNSGGGGTTPRAVGRVVALGQT 548

QY 682 VQASDLSLTGAETLNGLIQFDAALCPGDSGGPVVNGLGQVVGWNTAAS 729
Db 549 VQASDLSLTGAETLNGLIQFDAALCPGDSGGPVVNGLGQVVGWNTAAS 596

```

RESULT 4
 US-09-287-849-22
 ; Sequence 22, Application US/09287849
 ; Patent No. 6627198
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins
 ; TITLE OF INVENTION: and Their Uses
 ; FILE REFERENCE: 014058-009020US

```

; CURRENT APPLICATION NUMBER: US 09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/918,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

Query Match          53.4%; Score 1967.5; DB 4; Length 600;
Best Local Similarity 72.0%; Pred.No. 1e-135;
Matches 431; Conservative 19; Mismatches 82; Indels 67; Gaps 10;

QY      142  MVDFGALPPEINSARMYAGPGSASLVAAQMDSVASDLFSAASAFQSVVMGLTVGSWIG 201
      |||
DB      9   MVDFGALPPEINSARMYAGPGSASLVAAQMDSVASDLFSAASAFQSVVMGLTVGSWIG 68

QY      202  SSAGLVMVAASPVMWMSVTGAGAEITAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 261
      |||
DB      69  SSAGLVMVAASPVMWMSVTGAGAEITAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 128

QY      262  LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 321
      |||
DB      129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188

QY      322  LLEQAAVFEASTAAANQIMNNVPAQLQOLAQTOGTTPSSKLGGLWKTVPSPHRSPIIN 381
      |||
DB      189 LLEQAAVFEASTAAANQIMNNVPAQLQOLAQTOGTTPSSKLGGLWKTVPSPHRSPIIN 248

QY      382  MVSWMANNHSMTNSGVSWINTLTLSMLKGFAPAAARQAVQTAAQNGVRANSSLGSSLGSSG 441
      |||
DB      249 MVSWMANNHSMTNSGVSWINTLTLSMLKGFAPAAARQAVQTAAQNGVRANSSLGSSLGSSG 308

QY      442  LGGQVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGFQGMGLGPLV 501
      |||
DB      309 LGGQVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERGFQGMGLGPLV 368

QY      502  GQMGARAGGSLGVLPRPPPYWPHSPAAAGDTPA-----PALSQDR 543
      |||
DB      369 GQMGARAGGSLGVLPRPPPYWPHSPAAAGDTPVDVAVNTTCYGOVVAALNATDPGA 428

QY      544  FADFPALPLDPS-----AMVAQGFQVVVINTKLGYNNAVAGAGTGIVIDPN 589
      |||
DB      429 AAFQFNASPAQSVYLRNFLAAPPPQRAAMAAQL--QAVPEGAAQYITGLVESVAGSCN-----N 482

QY      590  GVVLTNNHVIAGATDINAFSGVSGQYGVVDVGVYDRQDVAVLQLRCAGGGLPSAAIGGV 649
      |||
DB      483 YELMTTNYQFG-----DVIDA-----HGMIRAQAAASLEAEHQAIKRVLDVLAAGDFWGA--GSV 533

QY      650  AVGPEPVVAMG-----NSGQGGTTPRAVGRVVVALGQTVOAOSLSLTGAE--ETLNG 697
      |||
DB      534 ACCEFTTOLGRNFOVIEQANAHGQ-----KVOAGNNNAQTDSAVGSSWATSG 583

```

RESULT 5
US-08-818-112-107
; Sequence 107, Application US/08818112
; Patent No. 6230969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.

APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedwick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELEPHONE: (206) 682-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-107

Query Match 52.7%; Score 1944; DB 3; Length 391;
Best Local Similarity 99.7%; Pred. No. 3e-134;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 142 MVDGALPPPEINSGARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 201
DB 1 MVDGALPPPEINSGARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60

QY 202 SSAGLVAAASPVYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
DB 61 SSAGLVAAASPVYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120

QY 262 LIATNLGQNTPAIVNEAEYGEWMAQDAAAVFGYAAATATATATATLLPPEEAPEMTSAGG 321
DB 121 LIATNLGQNTPAIVNEAEYGEWMAQDAAAVFGYAAATATATATLLPPEEAPEMTSAGG 180

QY 322 LLEQAAVEEASPTAAANQLMNNVPAALQQLAQPTGTTSSKLGGLWKTVPSPHSPISN 381
DB 181 LLEQAAVEEASPTAAANQLMNNVPAALQQLAQPTGTTSSKLGGLWKTVPSPHSPISN 240

QY 382 MVSMAHNMNTSGVSMNTNTLSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSG 441
DB 241 MVSMAHNMNTSGVSMNTNTLSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSG 300

QY 442 LGGGAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 501
DB 301 LGGGAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360

QY 502 GQNGARAGGGLSVLRVPRPYVPHSPAG 532
DB 361 GQNGARAGGGLSVLRVPRPYVPHSPAG 391

RESULT 6
US-08-818-111-102
Sequence 102, Application US/08818111
Patent No. 6338852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedwick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-102

Query Match 52.7%; Score 1944; DB 4; Length 391;
Best Local Similarity 99.7%; Pred. No. 3e-134;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 142 MVDGALPPPEINSGARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 201
DB 1 MVDGALPPPEINSGARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWGLTVGSMIG 60

QY 202 SSAGLVAAASPVYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
DB 61 SSAGLVAAASPVYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120

QY 262 LIATNLGQNTPAIVNEAEYGEWMAQDAAAVFGYAAATATATATLLPPEEAPEMTSAGG 321
DB 121 LIATNLGQNTPAIVNEAEYGEWMAQDAAAVFGYAAATATATATLLPPEEAPEMTSAGG 180

QY 322 LLEQAAVEEASPTAAANQLMNNVPAALQQLAQPTGTTSSKLGGLWKTVPSPHSPISN 381
DB 181 LLEQAAVEEASPTAAANQLMNNVPAALQQLAQPTGTTSSKLGGLWKTVPSPHSPISN 240

QY 382 MVSMAHNMNTSGVSMNTNTLSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSG 441
DB 241 MVSMAHNMNTSGVSMNTNTLSMLKGFAPAAAQAVQTAQAQNGVRAMSSLGSSG 300

QY 442 LGGGAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 501
DB 301 LGGGAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 360

Qy	502	QGMGARAGGG	LSGVL	RVPP	PRPY	MPHSPAAG	532
Db	361	QGMGARAGGG	LSGVL	RVPP	PRPY	MPHSPAAG	391

RESULT 7

US-09-056-556-107
; Sequence 107, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-107

Query Match	52.7%;	Score 1944;	DB 4;	Length 391;
Best Local Similarity	99.7%;	Pred. No. 3e-134;		
Matches 390;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	142	MVDFGALPPEINSRMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG	201	
Db	1	MVDFGALPPEINSRMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG	60	
Qy	202	SSAGLWVAASPPYVAMVSVTAGQAELETAQVRVAAAAYETAYGLTVPPPVAENRAELMI	261	
Db	61	SSAGLWVAASPPYVAMVSVTAGQAELETAQVRVAAAAYETAYGLTVPPPVAENRAELMI	120	
Qy	262	LIAENLIGQNTPAIAVNEAEYGEVWQAQAAAQFVYAAATATATATATLTPFEEAPEMTSAGG	321	
Db	121	LIAENLIGQNTPAIAVNEAEYGEVWQAQAAAQFVYAAATATATATATLTPFEEAPEMTSAGG	180	
Qy	322	LLEQAAVEEASDPTAAANQLMNNVPQALQQLAQQTQGTTPSSKLGGLWKTVSPHRSPISN	381	
Db	181	LLEQAAVEEASDPTAAANQLMNNVPQALQQLAQQTQGTTPSSKLGGLWKTVSPHRSPISN	240	
Qy	382	MVSMANNHMSMTNGSVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG	441	
Db	241	MVSMANNHMSMTNGSVSMNTLSSMLKGFAPAAAQAVQTAQNGVRAMSSLGSSLGSSG	300	
Qy	442	LGGVVAANLGRASAVGSLVSVPQAAWAAANQAVTPPAARALPLTSLTSAERGPQOMLGLPLV	501	
Db	301	LGGVVAANLGRASAVGSLVSVPQAAWAAANQAVTPPAARALPLTSLTSAERGPQOMLGLPLV	360	

QY	502	QGMGARAGGG	LSGVL	RVPP	PRPY	MPHSPAAG	532
4							
D6	361	QGMGARAGGG <th>LSGVL</th> <th>RVPP</th> <th>PRPY</th> <th>MPHSPAAG</th> <th>391</th>	LSGVL	RVPP	PRPY	MPHSPAAG	391

RESULT 8

```

US-09-072-596-102
; Sequence 102, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
; TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-09-072-596-102

```

	Query Match	52.7%	Score 1944;	DB 4;	Length 391;
	Best Local Similarity	99.7%;	Pred. No. 3e-134;		
	Matches 390;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0
Qy	142	MVDFGALPPPIINSARMYAGPGSLSLVAAAQWSDSVASDLFSAASAFQSVVMGLTVGSMIG	201		
Db	1	MVDFGALPPPIINSARMYAGPGSLSLVAAAQWSDSVASDLFSAASAFQSVVMGLTVGSMIG	60		
Qy	202	SSAGLIMVAAASPYVAMSVTGAQBELTAAQVRVAAAAAYETAYGLTVPPPVTAENRAELMI	261		
Db	61	SSAGLIMVAAASPYVAMSVTGAQBELTAAQVRVAAAAAYETAYGLTVPPPVTAENRAELMI	120		
Qy	262	LIAINLLGQNTPALAVNAEYGEVMAQDAAMFGYAAATATATATLPPFEEAPEMPTSGG	321		
Db	121	LIAINLLGQNTPALAVNAEYGEVMAQDAAMFGYAAATATATATLPPFEEAPEMPTSGG	180		
Qy	322	LLEQAAAAVEASDSTAANANQLMNNVPOALQQLAQPTQGTTPPSKLGGLMKWTVPHPHSPTSN	381		
Db	181	LLEQAAAAVEASDSTAANANQLMNNVPOALQQLAQPTQGTTPPSKLGGLMKWTVPHPHSPTSN	240		

QY 382 MYSMANNHMTNSGVSMTNTLSSMLKGFAPAPAAQAQVOTAAQNGVRAMSSLSLSSG 441
DB 241 MYSMANNHMTNSGVSMTNTLSSMLKGFAPAPAAQAQVOTAAQNGVRAMSSLSLSSG 300
QY 442 LGGVVAANLGRAASVGSLSVFOAANAANOAVTPAARALPLTSLTSAARPGQMLGLPV 501
DB 301 LGGVVAANLGRAASVGSLSVFOAANAANOAVTPAARALPLTSLTSAARPGQMLGLPV 360
QY 502 QOMGARAGGSLGVLVPRPPRYVPHSPAAG 532
DB 361 QOMGARAGGSLGVLVPRPPRYVPHSPAAG 391

RESULT 9
US-09-072-967-107
; Sequence 107, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; TYPE: amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
US-09-072-967-107

Query Match 52.7%; Score 1944; DB 4; Length 391;
Best Local Similarity 99.7%; Pred. No. 3e-134; 1; Indels 0; Gaps 0;
Matches 390; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 142 MVDGALPPEINSARMYAGFGSASLVAAQAQMDVSADLFSAAAFQSVVWGLTVGSMIG 201
DB 1 MVDGALPPEINSARMYAGFGSASLVAAQAQMDVSADLFSAAAFQSVVWGLTVGSMIG 60
QY 202 SSAGLVAAASPVVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 261
DB 61 SSAGLVAAASPVVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120

QY 262 LIATNLGONTFAIIVNEAEYGEWMAQDAAMFGYAAATATATATLTLFFEPAPEMTSAGG 321
DB 121 LIATNLGONTFAIIVNEAEYGEWMAQDAAMFGYAAATATATATLTLFFEPAPEMTSAGG 180
QY 322 LLEQAAAVVEASDTAAANQLMNNVFOALQOLAQOPTGGTTPSSKLGGLWKTVPKRSPTSN 381
DB 181 LLEQAAAVVEASDTAAANQLMNNVFOALQOLAQOPTGGTTPSSKLGGLWKTVPKRSPTSN 240
QY 382 MYSMANNHMTNSGVSMTNTLSSMLKGFAPAPAAQAQVOTAAQNGVRAMSSLSLSSG 441
DB 241 MYSMANNHMTNSGVSMTNTLSSMLKGFAPAPAAQAQVOTAAQNGVRAMSSLSLSSG 300
QY 442 LGGVVAANLGRAASVGSLSVFOAANAANOAVTPAARALPLTSLTSAARPGQMLGLPV 501
DB 301 LGGVVAANLGRAASVGSLSVFOAANAANOAVTPAARALPLTSLTSAARPGQMLGLPV 360
QY 502 QOMGARAGGSLGVLVPRPPRYVPHSPAAG 532
DB 361 QOMGARAGGSLGVLVPRPPRYVPHSPAAG 391

RESULT 10
US-08-818-112-111
; Sequence 111, Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,112
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.411C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-818-112-111

Query Match 44.8%; Score 1652.5; DB 3; Length 396;
Best Local Similarity 84.9%; Pred. No. 6.2e-113;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 142 MVDGALPPEINSARMYAGFGSASLVAAQAQMDVSADLFSAAAFQSVVWGLTVGSMIG 201
DB 1 MVDGALPPEINSARMYAGFGSASLVAAQAQMDVSADLFSAAAFQSVVWGLTVGSMIG 60

Db 1 VVDFGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTTGSGWIG 60
QY 202 SSAGLWVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 261
Db 61 SSAGLWVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
QY 262 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLPPPEAPMTSAGG 321
Db 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLPPPEAPMTSAGG 180
QY 322 LLEQAAVEEASDTAAANQLMNVPOALQLOAQPTQGTTPSSKLGGLWKTSPHRSPLSN 381
Db 181 LLEQAAVEEADTAAANQLMNVPOALQLOAQPTQGTTPSSKLGGLWKTSPHRSPLSN 240
QY 382 MVSMAHHMTNSGVSMTNTLSMLKGFAPAPAAQVOTAAQNGVRAMS-----LGSSL 437
Db 241 IVSMLNNHVSMTNSGVSMASTLHMLKGFAPAAA-QAVETAAQNGVQAMSSLGSLGSSL 299
QY 438 GSSGLGGVAAANLGRAASVGSLSVPOQAAWAAANQAVTPAARALPLTSLTSAABERGPGOMLG 497
Db 300 GSSGLGAGVAAANLGRAASVGSLSVPOQAAWAAANQAVTPAARALPLTSLTSAABERGPGOMLG 359
QY 498 GLPVGMQGARAG--GGLSGVLVPRPPRYVMPHSPAAG 532
Db 360 GLPLGLTNSGGGFGVSNALRMPPRAYVMPRVPRAAG 396

RESULT 11
US-08-818-111-106
; Sequence 106, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedrick, Thomas S.
; APPLICANT: Twardik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.41706
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-818-111-106

Query Match 44.8%; Score 1652.5; DB 4; Length 396;

Best Local Similarity 84.9%; Pred. No. 6.2e-113; Indels 7; Gaps 3;
Matches 337; Conservative 19; Mismatches 34;
QY 142 MVDFGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTTGSGWIG 201
Db 1 VVDFGALPPEINSARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTTGSGWIG 60
QY 202 SSAGLWVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 261
Db 61 SSAGLWVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 120
QY 262 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 321
Db 121 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180
QY 322 LLEQAAVEEASDTAAANQLMNVPOALQLOAQPTQGTTPSSKLGGLWKTSPHRSPLSN 381
Db 181 LLEQAAVEEADTAAANQLMNVPOALQLOAQPTQGTTPSSKLGGLWKTSPHRSPLSN 240
QY 382 MVSMAHHMTNSGVSMTNTLSMLKGFAPAPAAQVOTAAQNGVRAMS-----LGSSL 437
Db 241 IVSMLNNHVSMTNSGVSMASTLHMLKGFAPAAA-QAVETAAQNGVQAMSSLGSLGSSL 299
QY 438 GSSGLGGVAAANLGRAASVGSLSVPOQAAWAAANQAVTPAARALPLTSLTSAABERGPGOMLG 497
Db 300 GSSGLGAGVAAANLGRAASVGSLSVPOQAAWAAANQAVTPAARALPLTSLTSAABERGPGOMLG 359
QY 498 GLPVGMQGARAG--GGLSGVLVPRPPRYVMPHSPAAG 532
Db 360 GLPLGLTNSGGGFGVSNALRMPPRAYVMPRVPRAAG 396

RESULT 12
US-09-056-556-111
; Sequence 111, Application US/09056556
; Patent No. 6350456
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
; NUMBER OF SEQUENCES: 241
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 111:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-09-056-556-111

TREATME

Query Match	44.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity	84.9%; Pred. No. 6.2e-113;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;	
QY	142 MVDGALPEINSARMYAGPGSASLVAQAQWDSVASDLFSAASAFQSVVWGLTGVSGWIG 201
Db	1 VVDGALPEINSARMYAGPGSASLVAQAQWDSVASDLFSAASAFQSVVWGLTGVSGWIG 60
QY	202 SSAGLMVAAAAPYVAMVSVTAGQAEITAAQVRVAAAAAYETAYGLTVPFPVIAENRAELMI 261
Db	61 SSAGLMVAAAAPYVAMVSVTAGQAEITAAQVRVAAAAAYETAYGLTVPFPVIAENRAELMI 120
QY	262 LIATNLLGONTPATVAVNEAEYGEWQAQDAAMEGYAAATATATATLTFEAPBEMTSAGG 321
Db	121 LIATNLLGONTPATVAVNEAEYGEWQAQDAAMEGYAAATATATLTFEAPBEMTIPGG 180
QY	322 LLEQAAVVEEASDTAAANQLMNVNQALQOLAQPTQGTTPSSKLGGLMKTVPSPHRSPISN 381
Db	181 LLEQAVAVEEADITAAANQLMNVNQALQOLAQPTKSIWPPDQSELWKATSPHLSPLSN 240
QY	382 MYSMANHVMSTNSGVSNWTTLSMLKGFAPAAAQAVOTAAQNGVRAMSS-----LGSLL 437
Db	241 IVSMLNHVMSNTNSGVSNMSTLSMLKGFAPAAA-QAVETAAQNGVQVAMSSLGSLGSSLL 299
QY	438 GSSGLGGGVAANTLGRAASVCSLSVPQAAAAANQAVTPAARALPLTSLTSAACERPGOMLG 497
Db	300 GSSGLGGGVAANTLGRAASVCSLSVPQAAAAANQAVTPAARALPLTSLTSAATQAPGMLG 359
QY	498 GLPVQMGARAG--GGLSGVLRVPPRPVYMPHSPAAG 532
Db	360 GLPLGOLTNSSGGGGVSNALRMEPRAYVYVPRVPAAG 396

RESULT 13
 US-09-072-596-106
 ; Sequence 106, Application US/09072596
 ; Patent No. 6458366
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonia
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Hendrickson, Ronald C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 ; TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 350
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED and BERRY LLP
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072,596
 ; FILING DATE: 05-MAY-1998
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.417C9
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 106:

RESULT 14
US-09-072-967-111
; Sequence 111, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 355
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Avenue, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,967
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 111:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-09-072-967-111

Query Match 44.8%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 6.2e-113; Indels 7; Gaps 3;
Matches 337; Conservative 19; Mismatches 34

Qy 142 MVDGALPPEINARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGWSWIG 201
Db 1 VVDFGALPPEINARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGWSWIG 60
Qy 202 SSAGLMVAASPYVAMSVTAGOAEITAAQVRVAAAAYETAYGLTPPPVIAENRAELMI 261
Db 61 SSAGLMVAASPYVAMSVTAGOAEITAAQVRVAAAAYETAYGLTPPPVIAENRAELMI 120
Qy 262 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLPPFEAPENTSAGG 321
Db 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATLPPFEAPENTSAGG 180
Qy 322 LLEQAAVEEASDTAAANQLMNNVPOALQLOAQPTGTPSKLGGWKTVSPHRSPISN 381
Db 181 LLEQAAVEEASDTAAANQLMNNVPOALQLOAQPTGTPSKLGGWKTVSPHRSPISN 240
Qy 382 MYSMANNHMTNSGVSMNTLSSMLKGFAPAPAAARQAVTAAQNGVRAMSS---LGSSL 437
Db 241 IVSMNNHMTNSGVSMNTLSSMLKGFAPAPAAARQAVTAAQNGVRAMSS---LGSSL 299
Qy 438 GSSGLGGVAAANLGRAASVGSLSVFOAWAANAQAVTPAARALPLTSLTSAABERGQMLG 497
Db 300 GSSGLGGVAAANLGRAASVGSLSVFOAWAANAQAVTPAARALPLTSLTSAABERGQMLG 359
Qy 498 GLPVQMGARAG--CGLSGLVLRPPRYMHPSPAG 532
Db 360 GLPLGOLTNSSGGFGGVSNALRMPPRAYMHPVPAAG 396

RESULT 15
US-08-818-112-109
Sequence 109, Application US/08818112
Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

Search completed: June 30, 2004, 16:55:22
Job time : 24.1761 secs

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818,112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 109:
SEQUENCE CHARACTERISTICS:
LENGTH: 359 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-818-112-109

Query Match 40.3%; Score 1486.5; DB 3; Length 359;
Best Local Similarity 84.2%; Pred. No. 7.3e-101;
Matches 303; Conservative 16; Mismatches 36; Indels 5; Gaps 2;

Qy 142 MVDGALPPEINARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGWSWIG 201
Db 1 VVDFGALPPEINARMYAGPGSASLVAAAKWDSVASDLFSAASAFQSVVWGLTVGWSWIG 60
Qy 202 SSAGLMVAASPYVAMSVTAGOAEITAAQVRVAAAAYETAYGLTPPPVIAENRAELMI 261
Db 61 SSAGLMVAASPYVAMSVTAGOAEITAAQVRVAAAAYETAYGLTPPPVIAENRAELMT 120
Qy 262 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATLPPFEAPENTSAGG 321
Db 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATLPPFEAPENTSAGG 180
Qy 322 LLEQAAVEEASDTAAANQLMNNVPOALQLOAQPTGTPSKLGGWKTVSPHRSPISN 381
Db 181 LLEQAAVEEASDTAAANQLMNNVPOALQLOAQPTGTPSKLGGWKTVSPHRSPISN 240
Qy 382 MYSMANNHMTNSGVSMNTLSSMLKGFAPAPAAARQAVTAAQNGVRAMSS---LGSSL 437
Db 241 VSSIANNHMTNSGVSMNTLSSMLKGFAPAPAAARQAVTAAQNGVRAMSS---LGSSL 299
Qy 438 GSSGLGGVAAANLGRAASVGSLSVFOAWAANAQAVTPAARALPLTSLTSAABERGQMLG 497
Db 300 GSSGLGGVAAANLGRAASVGSLSVFOAWAANAQAVTPAARALPLTSLTSAABERGQMLG 359

The first part of the document discusses the importance of maintaining accurate records of all transactions. It emphasizes that proper record-keeping is essential for the integrity of the financial system and for the ability to detect and prevent fraud. The document also outlines the responsibilities of individuals involved in the process, including the need for transparency and accountability.

In addition, the document highlights the role of technology in improving record-keeping and data management. It suggests that the use of digital tools can help to streamline the process and reduce the risk of errors. The document also mentions the importance of regular audits and reviews to ensure that the system is functioning effectively.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:52:58 ; Search time 56.8458 seconds
(without alignments)
3625.462 Million cell updates/sec

Title: US-09-597-796C-12

Perfect score: 3686

Sequence: 1 MHHHHTAASDNFQLSQGG.....SGGPVNVGLGVGMNTAAS 729

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1166195 seqs, 282705291 residues

Total number of hits satisfying chosen parameters: 1166195

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3686	100.0	729	9	US-09-287-849-2
2	3686	100.0	729	12	US-09-886-349A-16
3	3686	100.0	729	14	US-10-359-460-2
4	3686	100.0	729	14	US-10-098-732A-16
5	3686	100.0	729	15	US-10-359-459-2
6	3680	99.8	729	15	US-10-369-983-21
7	3680	99.8	813	15	US-10-369-983-15
8	3680	99.8	825	15	US-10-369-983-14
9	3680	99.8	875	15	US-10-369-983-13
10	3680	99.8	930	14	US-10-098-732A-65
11	3680	99.8	930	15	US-10-369-983-12
12	3680	99.8	1016	15	US-10-369-983-18
13	3680	99.8	1022	15	US-10-369-983-17
14	3680	99.8	1154	15	US-10-369-983-16
15	3677	99.8	729	12	US-09-886-349A-18

16	3677	99.8	729	14	US-10-098-732A-18	Sequence 18, Appl
17	3677	99.8	729	15	US-10-369-983-22	Sequence 22, Appl
18	2941	79.8	596	9	US-09-287-849-26	Sequence 26, Appl
19	2941	79.8	596	12	US-09-886-349A-20	Sequence 20, Appl
20	2941	79.8	596	14	US-10-359-460-25	Sequence 26, Appl
21	2941	79.8	596	14	US-10-098-732A-20	Sequence 20, Appl
22	2637	71.5	1010	15	US-10-369-983-4	Sequence 4, Appl
23	2631	71.4	723	15	US-10-369-983-2	Sequence 22, Appl
24	1967.5	53.4	600	9	US-09-287-849-22	Sequence 22, Appl
25	1967.5	53.4	600	14	US-10-359-460-22	Sequence 22, Appl
26	1944	52.7	391	12	US-09-886-349A-14	Sequence 14, Appl
27	1944	52.7	391	14	US-10-193-002-102	Sequence 102, App
28	1944	52.7	391	14	US-10-084-843-107	Sequence 107, App
29	1944	52.7	391	14	US-10-098-732A-14	Sequence 14, Appl
30	1929	52.3	391	12	US-09-872-186-8	Sequence 8, Appl
31	1652.5	44.8	396	14	US-10-193-002-106	Sequence 106, App
32	1652.5	44.8	396	14	US-10-084-843-111	Sequence 111, App
33	1583	42.9	393	12	US-10-282-132A-62455	Sequence 62455, A
34	1583	42.9	393	12	US-10-282-132A-64892	Sequence 64892, A
35	1486.5	40.3	359	14	US-10-193-002-104	Sequence 104, App
36	1486.5	40.3	359	14	US-10-084-843-109	Sequence 109, App
37	1184	32.1	358	9	US-09-287-849-8	Sequence 8, Appl
38	1184	32.1	358	14	US-10-359-460-8	Sequence 8, Appl
39	1182	32.1	263	12	US-09-886-349A-12	Sequence 12, Appl
40	1182	32.1	263	14	US-10-193-002-92	Sequence 92, Appl
41	1182	32.1	263	14	US-10-084-843-91	Sequence 91, Appl
42	1182	32.1	263	14	US-10-098-732A-12	Sequence 12, Appl
43	989	26.8	355	9	US-09-712-363-161	Sequence 161, App
44	987	26.8	330	12	US-09-886-349A-4	Sequence 4, Appl
45	987	26.8	330	14	US-10-098-732A-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1

US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Protiens of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-2

Query Match 100.0%; Score 3686; DB 9; Length 729;
Best Local Similarity 100.0%; Pred. No. 7.5e-244;

Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEHHHHTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPTHVHIGTAFGLGVVD 60
DB 1 MEHHHHTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPTHVHIGTAFGLGVVD 60

QY 61 NNGGARVQVRVGSAPASISIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
DB 61 NNGGARVQVRVGSAPASISIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120

QY 121 TKSGETTGNVTLAGPPAEFMDVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
DB 121 TKSGETTGNVTLAGPPAEFMDVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSSWIGSAGLMVAASPYVAMSVTACQAEILTAQVRAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSSWIGSAGLMVAASPYVAMSVTACQAEILTAQVRAAAAYE 240

QY 241 TAYGLTVPPIAENRAELMILITATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPIAENRAELMILITATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300

QY 301 ATATATLLPPEEAPMTSAGLLEQAAAVVEASDTAAANQLMNVPOALQQLAQTQGT 360
DB 301 ATATATLLPPEEAPMTSAGLLEQAAAVVEASDTAAANQLMNVPOALQQLAQTQGT 360

QY 361 PSSKLGKLTWTSVPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
DB 361 PSSKLGKLTWTSVPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQ 420

QY 421 TAAQNGVRAMSSLGSSIGSSGLGGVAANLGRAASVGSLSVPOAWAANCAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSIGSSGLGGVAANLGRAASVGSLSVPOAWAANCAVTPAARALP 480

QY 481 LTSLSAARGPGQMLGGLPVGQMGARAGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
DB 481 LTSLSAARGPGQMLGGLPVGQMGARAGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540

QY 541 QDRFADFPALPLDPSAMVAQVGPVNVNINIKLGYNNAVGAGTGIVDPNGVLTNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPVNVNINIKLGYNNAVGAGTGIVDPNGVLTNNHVA 600

QY 601 GATDINAFSVSGQTYGVDVVGVDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 660
DB 601 GATDINAFSVSGQTYGVDVVGVDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 660

QY 661 SGGQGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLQ 720
DB 661 SGGQGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLQ 720

QY 721 VVGWNTAAS 729
DB 721 VVGWNTAAS 729

RESULT 2

US-09-886-349A-16
; Sequence 16, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50

; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 16
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein MTB72F (Ra12-TbH9-Ra35 or MTB32-MTB39
; OTHER INFORMATION: fusion)
US-09-886-349A-16

Query Match 100.0%; Score 3686; DB 12; Length 729;
Best Local Similarity 100.0%; Pred. No. 7.5e-244;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MEHHHHTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPTHVHIGTAFGLGVVD 60
DB 1 MEHHHHTAASDNFQLSQGGGFAIPIGQAMATAGQIRSGGSPTHVHIGTAFGLGVVD 60

QY 61 NNGGARVQVRVGSAPASISIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
DB 61 NNGGARVQVRVGSAPASISIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120

QY 121 TKSGETTGNVTLAGPPAEFMDVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
DB 121 TKSGETTGNVTLAGPPAEFMDVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSSWIGSAGLMVAASPYVAMSVTACQAEILTAQVRAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSSWIGSAGLMVAASPYVAMSVTACQAEILTAQVRAAAAYE 240

QY 241 TAYGLTVPPIAENRAELMILITATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPIAENRAELMILITATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300

QY 301 ATATATLLPPEEAPMTSAGLLEQAAAVVEASDTAAANQLMNVPOALQQLAQTQGT 360
DB 301 ATATATLLPPEEAPMTSAGLLEQAAAVVEASDTAAANQLMNVPOALQQLAQTQGT 360

QY 361 PSSKLGKLTWTSVPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
DB 361 PSSKLGKLTWTSVPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQ 420

QY 421 TAAQNGVRAMSSLGSSIGSSGLGGVAANLGRAASVGSLSVPOAWAANCAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSIGSSGLGGVAANLGRAASVGSLSVPOAWAANCAVTPAARALP 480

QY 481 LTSLSAARGPGQMLGGLPVGQMGARAGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
DB 481 LTSLSAARGPGQMLGGLPVGQMGARAGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540

QY 541 QDRFADFPALPLDPSAMVAQVGPVNVNINIKLGYNNAVGAGTGIVDPNGVLTNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPVNVNINIKLGYNNAVGAGTGIVDPNGVLTNNHVA 600

QY 601 GATDINAFSVSGQTYGVDVVGVDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 660
DB 601 GATDINAFSVSGQTYGVDVVGVDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGN 660

QY 661 SGGQGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLQ 720
DB 661 SGGQGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLQ 720

QY 721 VVGWNTAAS 729
DB 721 VVGWNTAAS 729

RESULT 3

US-10-359-460-2
; Sequence 2, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:

APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, David C.
APPLICANT: Alderson, Mark
APPLICANT: Campos-Neto, Antonio
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
TITLE OF INVENTION: and Their Uses
FILE REFERENCE: 014058-009020US
CURRENT APPLICATION NUMBER: US/10/359,460
CURRENT FILING DATE: 2003-02-05
PRIOR APPLICATION NUMBER: US/09/287,849
PRIOR FILING DATE: 1999-04-07
PRIOR APPLICATION NUMBER: US 08/818,112
PRIOR FILING DATE: 1997-03-13
PRIOR APPLICATION NUMBER: US 08/942,578
PRIOR FILING DATE: 1997-10-01
PRIOR APPLICATION NUMBER: US 09/025,197
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 09/056,556
PRIOR FILING DATE: 1998-04-07
PRIOR APPLICATION NUMBER: US 09/223,040
PRIOR FILING DATE: 1998-12-30
NUMBER OF SEQ ID NOS: 46
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 2
LENGTH: 729
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-10-359-460-2

Query Match 100.0%; Score 3686; DB 14; Length 729;
Best Local Similarity 100.0%; Pred. No. 7.5e-244;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHTAASDNFQSQGGGFAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQSQGGGFAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
Qy 61 NNGGARVQVVGSAASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
Db 61 NNGGARVQVVGSAASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
Qy 121 TKSGGTGTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
Db 121 TKSGGTGTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
Qy 181 FSAASAFQSVVWGLTVGSGWIGSAGLMVAASPYVAMSVTAGQAEITAAQVRAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWIGSAGLMVAASPYVAMSVTAGQAEITAAQVRAAAAYE 240
Qy 241 TAYGLTVPPVIAENRAELMILITNLLGQNTPAIAVNEAEGEMWADAAAFYAAAT 300
Db 241 TAYGLTVPPVIAENRAELMILITNLLGQNTPAIAVNEAEGEMWADAAAFYAAAT 300
Qy 301 ATATATLLPFEEAPEMTSAGGLEQAAAVEEASDTAAANQLMNNVPQALQQAQPTQGT 360
Db 301 ATATATLLPFEEAPEMTSAGGLEQAAAVEEASDTAAANQLMNNVPQALQQAQPTQGT 360
Qy 361 PSSKLGGLWKTVPSPHRSPTSNMVMANNHMTNSGVSMNTLSSMLKGFAPAAARQAV 420
Db 361 PSSKLGGLWKTVPSPHRSPTSNMVMANNHMTNSGVSMNTLSSMLKGFAPAAARQAV 420
Qy 421 TAAQNGVRAMSSIGSLGSGGVAANLGRAASVGSISVPQAAVAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSIGSLGSGGVAANLGRAASVGSISVPQAAVAANQAVTPAARALP 480
Qy 481 L7SLTSAERPGCOMLGGPLPVQMGARAGGGLSGVLVPPRPYVPHSPAGDIAAPPALS 540
Db 481 L7SLTSAERPGCOMLGGPLPVQMGARAGGGLSGVLVPPRPYVPHSPAGDIAAPPALS 540

Qy 541 ODRFADFPALDPSAMVAQVGPVNVNTKLGYNNAVGAGTGIVDPNGVLTNNHVIA 600
Db 541 ODRFADFPALDPSAMVAQVGPVNVNTKLGYNNAVGAGTGIVDPNGVLTNNHVIA 600
Qy 601 GATDINAFSVGSGQTYGVVDVVDYDRTQDVAVLQRGAGLPSAAIGGVAVGPPVAMGN 660
Db 601 GATDINAFSVGSGQTYGVVDVVDYDRTQDVAVLQRGAGLPSAAIGGVAVGPPVAMGN 660
Qy 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIOPGDSGGPVVNLGQ 720
Db 661 SGGGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIOPGDSGGPVVNLGQ 720
Qy 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

RESULT 4
US-10-098-732A-16
Sequence 16, Application US/10098732A
Publication No. US20030175294A1
GENERAL INFORMATION:
APPLICANT: Skeiky, Yasir
APPLICANT: Brannon, Mark
APPLICANT: Guiderian, Jeffrey
APPLICANT: Corixa Corporation
TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
TITLE OF INVENTION: Leishmania Antigen
FILE REFERENCE: 014058-012010US
CURRENT APPLICATION NUMBER: US/10/098,732A
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: US 60/275,837
PRIOR FILING DATE: 2001-03-13
NUMBER OF SEQ ID NOS: 80
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 16
LENGTH: 729
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
OTHER INFORMATION: Protein MTB72F (Ra12-TbH9-Ra35 or MTB32-MTB39
OTHER INFORMATION: fusion)
US-10-098-732A-16

Query Match 100.0%; Score 3686; DB 14; Length 729;
Best Local Similarity 100.0%; Pred. No. 7.5e-244;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MHHHHHTAASDNFQSQGGGFAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQSQGGGFAIPGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
Qy 61 NNGGARVQVVGSAASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
Db 61 NNGGARVQVVGSAASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
Qy 121 TKSGGTGTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
Db 121 TKSGGTGTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
Qy 181 FSAASAFQSVVWGLTVGSGWIGSAGLMVAASPYVAMSVTAGQAEITAAQVRAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWIGSAGLMVAASPYVAMSVTAGQAEITAAQVRAAAAYE 240
Qy 241 TAYGLTVPPVIAENRAELMILITNLLGQNTPAIAVNEAEGEMWADAAAFYAAAT 300
Db 241 TAYGLTVPPVIAENRAELMILITNLLGQNTPAIAVNEAEGEMWADAAAFYAAAT 300
Qy 301 ATATATLLPFEEAPEMTSAGGLEQAAAVEEASDTAAANQLMNNVPQALQQAQPTQGT 360
Db 301 ATATATLLPFEEAPEMTSAGGLEQAAAVEEASDTAAANQLMNNVPQALQQAQPTQGT 360


```
Db 61 NNGNGARVQVVGSAASLGIISTGDVITAVDGAIPINSATAMADALNGHHPGCVISVTWQ 120
QY 121 TSKGGTRTGNVTLAGSPPAEFWDFGALPPEINSAARMYAGPGSASLVAAQWMDSVASDL 180
Db 121 TSKGGTRTGNVTLAGSPPAEFWDFGALPPEINSAARMYAGPGSASLVAAQWMDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWISGAGLMVAASAPYVAMSVTAGOAEHTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWISGAGLMVAASAPYVAMSVTAGOAEHTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVIAENRAELMILIAIATNLGONTPAIAVNEAEYGEWMAQDAAMFYAAAT 300
Db 241 TAYGLTVPPVIAENRAELMILIAIATNLGONTPAIAVNEAEYGEWMAQDAAMFYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGGLLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Db 301 ATATATLLPFEAPEMTSAGGLLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSSGAGLVGQMGARAGGGLSGVLRVPPYVMPHSPAAGDIAPPALS 540
Db 421 TAAQNGVRAMSSLGSSLGSSGAGLVGQMGARAGGGLSGVLRVPPYVMPHSPAAGDIAPPALS 540
QY 541 QDRPADFPALPLDPSAMVAQVGPVNVNINIKLGYNNVAVGAGTGIVDPNGVVLTNHNVIA 600
Db 541 QDRPADFPALPLDPSAMVAQVGPVNVNINIKLGYNNVAVGAGTGIVDPNGVVLTNHNVIA 600
QY 601 GATDINAFSVGSGQTYGVDVGVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
Db 601 GATDINAFSVGSGQTYGVDVGVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
QY 721 VVGWNTAAS 729
Db 721 VVGWNTAAS 729
```

RESULT 7

```
US-10-369-983-15
; Sequence 15, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 15
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB81F (MTB72F-DFV)
US-10-369-983-15
```

```
Query Match 99.8%; Score 3680; DB 15; Length 813;
Best Local Similarity 99.9%; Pred. No. 2.2e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MHHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPPTVHIGTAFILGLGVD 60
Db 1 MHHHHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSPPTVHIGTAFILGLGVD 60
QY 61 NNGNGARVQVVGSAASLGIISTGDVITAVDGAIPINSATAMADALNGHHPGCVISVTWQ 120
Db 61 NNGNGARVQVVGSAASLGIISTGDVITAVDGAIPINSATAMADALNGHHPGCVISVTWQ 120
QY 121 TSKGGTRTGNVTLAGSPPAEFWDFGALPPEINSAARMYAGPGSASLVAAQWMDSVASDL 180
Db 121 TSKGGTRTGNVTLAGSPPAEFWDFGALPPEINSAARMYAGPGSASLVAAQWMDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWISGAGLMVAASAPYVAMSVTAGOAEHTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWISGAGLMVAASAPYVAMSVTAGOAEHTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVIAENRAELMILIAIATNLGONTPAIAVNEAEYGEWMAQDAAMFYAAAT 300
Db 241 TAYGLTVPPVIAENRAELMILIAIATNLGONTPAIAVNEAEYGEWMAQDAAMFYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGGLLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Db 301 ATATATLLPFEAPEMTSAGGLLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSSGAGLVGQMGARAGGGLSGVLRVPPYVMPHSPAAGDIAPPALS 480
Db 421 TAAQNGVRAMSSLGSSLGSSGAGLVGQMGARAGGGLSGVLRVPPYVMPHSPAAGDIAPPALS 480
QY 481 LTSLSAERGPQMLGGLPVGQMGARAGGGLSGVLRVPPYVMPHSPAAGDIAPPALS 540
Db 481 LTSLSAERGPQMLGGLPVGQMGARAGGGLSGVLRVPPYVMPHSPAAGDIAPPALS 540
QY 541 QDRPADFPALPLDPSAMVAQVGPVNVNINIKLGYNNVAVGAGTGIVDPNGVVLTNHNVIA 600
Db 541 QDRPADFPALPLDPSAMVAQVGPVNVNINIKLGYNNVAVGAGTGIVDPNGVVLTNHNVIA 600
QY 601 GATDINAFSVGSGQTYGVDVGVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
Db 601 GATDINAFSVGSGQTYGVDVGVYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLGTGABETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
Db 661 SGGGGTTPRAVGRVVALGQTVQASDSLGTGABETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
QY 721 VVGWNTAAS 729
Db 721 VVGWNTAAS 729
```

RESULT 8

```
US-10-369-983-14
; Sequence 14, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
```

```
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 825
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB83F (MTB72F-MTI)
US-10-369-983-14

Query Match      99.8%; Score 3680; DB 15; Length 825;
Best Local Similarity 99.8%; Pred. No. 2.3e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGQGFPAIGQAMATAGQIRSGGSPTHVIGTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQLSQGGQGFPAIGQAMATAGQIRSGGSPTHVIGTAFGLGVVD 60
QY 61 NNGNGARVQVRVGSAPAAASLGISTGDIVITAVDGPINSATAMADALNGHHPGDIVISVTWQ 120
DB 61 NNGNGARVQVRVGSAPAAASLGISTGDIVITAVDGPINSATAMADALNGHHPGDIVISVTWQ 120
QY 121 TKSGETRTGNVTLLAEGPPAEFMDFGALPPEINSARYAGPGSASLVAAQAQWDSVASDL 180
DB 121 TKSGETRTGNVTLLAEGPPAEFMDFGALPPEINSARYAGPGSASLVAAQAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMSVTAGQAELETAQVRVAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMSVTAGQAELETAQVRVAAAAYE 240
QY 241 TAYGLTVPPPIAENRAELMILITATNLGONTPAIAVNEAEYGEWMAQAAAFVGAAT 300
DB 241 TAYGLTVPPPIAENRAELMILITATNLGONTPAIAVNEAEYGEWMAQAAAFVGAAT 300
QY 301 ATATATLLPPEEAPMTSAGGLLEQAAAEEASDTAAANQLMNVPOALQQAQPTQGT 360
DB 301 ATATATLLPPEEAPMTSAGGLLEQAAAEEASDTAAANQLMNVPOALQQAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHMSWNTNSGVSMNTLSSMLKGFAPAARAAQVQ 420
DB 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHMSWNTNSGVSMNTLSSMLKGFAPAARAAQVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPQAAANQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPQAAANQAVTPAARALP 480
QY 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSGVLRVPPRYMPHSPAAGDIAPPALS 540
DB 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSGVLRVPPRYMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGVQVNNINTKLYNNNAVAGTGIVDPNGVLTNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGVQVNNINTKLYNNNAVAGTGIVDPNGVLTNNHVA 600
QY 601 GATDINAFSVSGGQTYGVDVVDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMGN 660
DB 601 GATDINAFSVSGGQTYGVDVVDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
DB 661 SGGGGTTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
QY 721 VVGNTAAS 729
DB 721 VVGNTAAS 729
```

RESULT 9

```
US-10-369-983-13
; Sequence 13, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIORITY FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIORITY FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 875
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB89F (MTB72F-Erd14)
US-10-369-983-13
```

```
Query Match      99.8%; Score 3680; DB 15; Length 875;
Best Local Similarity 99.8%; Pred. No. 2.4e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGQGFPAIGQAMATAGQIRSGGSPTHVIGTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQLSQGGQGFPAIGQAMATAGQIRSGGSPTHVIGTAFGLGVVD 60
QY 61 NNGNGARVQVRVGSAPAAASLGISTGDIVITAVDGPINSATAMADALNGHHPGDIVISVTWQ 120
DB 61 NNGNGARVQVRVGSAPAAASLGISTGDIVITAVDGPINSATAMADALNGHHPGDIVISVTWQ 120
QY 121 TKSGETRTGNVTLLAEGPPAEFMDFGALPPEINSARYAGPGSASLVAAQAQWDSVASDL 180
DB 121 TKSGETRTGNVTLLAEGPPAEFMDFGALPPEINSARYAGPGSASLVAAQAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMSVTAGQAELETAQVRVAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSSAGLMVAASPYVAMSVTAGQAELETAQVRVAAAAYE 240
QY 241 TAYGLTVPPPIAENRAELMILITATNLGONTPAIAVNEAEYGEWMAQAAAFVGAAT 300
DB 241 TAYGLTVPPPIAENRAELMILITATNLGONTPAIAVNEAEYGEWMAQAAAFVGAAT 300
QY 301 ATATATLLPPEEAPMTSAGGLLEQAAAEEASDTAAANQLMNVPOALQQAQPTQGT 360
DB 301 ATATATLLPPEEAPMTSAGGLLEQAAAEEASDTAAANQLMNVPOALQQAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHMSWNTNSGVSMNTLSSMLKGFAPAARAAQVQ 420
DB 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHMSWNTNSGVSMNTLSSMLKGFAPAARAAQVQ 420
QY 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPQAAANQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPQAAANQAVTPAARALP 480
QY 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSGVLRVPPRYMPHSPAAGDIAPPALS 540
DB 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSGVLRVPPRYMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGVQVNNINTKLYNNNAVAGTGIVDPNGVLTNNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGVQVNNINTKLYNNNAVAGTGIVDPNGVLTNNHVA 600
QY 601 GATDINAFSVSGGQTYGVDVVDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMGN 660
DB 601 GATDINAFSVSGGQTYGVDVVDRTQDVAVLQIRGAGGLPSAAIGGVAVGEPVVMGN 660
QY 661 SGGGGTTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
DB 661 SGGGGTTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNLGQ 720
```

```
QY 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

RESULT 10
US-10-098-732A-65
; Sequence 65, Application US/10098732A
; Publication NO. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Guderian, Jeff
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; FILE REFERENCE: Leishmania Antigen
; CURRENT APPLICATION NUMBER: US/10/098,732A
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 65
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72F-MAPS
; OTHER INFORMATION: (r95f) fusion construct, TB MTB72F (Ra12-TbH9-Ra35)
; OTHER INFORMATION: linked to Leishmania thiol-specific antioxidant
; OTHER INFORMATION: (TSA or MAPS)
US-10-098-732A-65

Query Match 99.8%; Score 3680; DB 14; Length 930;
Best Local Similarity 99.9%; Pred. No. 2.6e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60

QY 61 NNGNGARVQVRVGSAPAAASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
Db 61 NNGNGARVQVRVGSAPAAASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120

QY 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
Db 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSGWIGSSAGLMVAASPYVAMWSVTAGQELTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWIGSSAGLMVAASPYVAMWSVTAGQELTAAQVRVAAAAYE 240

QY 241 TAYGLTVPPVIAENRAELMILIAATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVIAENRAELMILIAATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300

QY 301 ATATATLLPFEAPENTTSAGGLEQAAAEEASDTAAANQLMNNVPOALQQLAQPTQGT 360
Db 301 ATATATLLPFEAPENTTSAGGLEQAAAEEASDTAAANQLMNNVPOALQQLAQPTQGT 360

QY 361 PSSKLGLWKTVPSPHRSPISNVMVMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
Db 361 PSSKLGLWKTVPSPHRSPISNVMVMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420

QY 421 TAAQNGVRAMSSIGSSIGSLGGGVAANI GRAASVGSLSVPCAWAAANQAVTPPAARALP 480
Db 421 TAAQNGVRAMSSIGSSIGSLGGGVAANI GRAASVGSLSVPCAWAAANQAVTPPAARALP 480

QY 481 LTSLSAAERGPQMLGGLPVGQMGARAGGSLGVLVRPPRVPYMPHSPAAGDIAPPALS 540
Db 481 LTSLSAAERGPQMLGGLPVGQMGARAGGSLGVLVRPPRVPYMPHSPAAGDIAPPALS 540
```

```
QY 541 QDRFADFPALPLDPSAMVAQVGPOVNVNINTKLYNNNAVAGTGIVIDPNGVVLTNHVLIA 600
Db 541 QDRFADFPALPLDPSAMVAQVGPOVNVNINTKLYNNNAVAGTGIVIDPNGVVLTNHVLIA 600

QY 601 GATDINAFSVGSGQTYGVYDVTQDVAVLQLRGAGGLPSAAIGGVAVGSPVVMAGN 660
Db 601 GATDINAFSVGSGQTYGVYDVTQDVAVLQLRGAGGLPSAAIGGVAVGSPVVMAGN 660

QY 661 SGGGGTTPRAVPRVVALGQTVQASDSLTCABETLNGLIQFDAAIQBDSGGPVVNLGQ 720
Db 661 SGGGGTTPRAVPRVVALGQTVQASDSLTCABETLNGLIQFDAAIQBDSGGPVVNLGQ 720

QY 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

RESULT 11
US-10-369-983-12
; Sequence 12, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:fusion protein
; OTHER INFORMATION: R95F (MTB72F-MAPS)
US-10-369-983-12

Query Match 99.8%; Score 3680; DB 15; Length 930;
Best Local Similarity 99.9%; Pred. No. 2.6e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQSQGGGFAIPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGVVD 60

QY 61 NNGNGARVQVRVGSAPAAASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120
Db 61 NNGNGARVQVRVGSAPAAASLGISTGDIVITAVDGPINSATAMADALNGHHPGDVSVTWQ 120

QY 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180
Db 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQMWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSGWIGSSAGLMVAASPYVAMWSVTAGQELTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSGWIGSSAGLMVAASPYVAMWSVTAGQELTAAQVRVAAAAYE 240

QY 241 TAYGLTVPPVIAENRAELMILIAATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVIAENRAELMILIAATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300

QY 301 ATATATLLPFEAPENTTSAGGLEQAAAEEASDTAAANQLMNNVPOALQQLAQPTQGT 360
Db 301 ATATATLLPFEAPENTTSAGGLEQAAAEEASDTAAANQLMNNVPOALQQLAQPTQGT 360

QY 361 PSSKLGLWKTVPSPHRSPISNVMVMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
Db 361 PSSKLGLWKTVPSPHRSPISNVMVMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
```

Db 361 PSSKLGGLWKTVPSPHRSPIINMWSMANNHSMNTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
Qy 421 TAAQNGVRAMSSLSGSSLSGSSGGLGGVAANTLGRAASVGSLSVPQAWAANAQVTPAARALP 480
Db 421 TAAQNGVRAMSSLSGSSLSGSSGGLGGVAANTLGRAASVGSLSVPQAWAANAQVTPAARALP 480
Qy 481 LTSLSAAERPGQMLGGLPVQMGARAGGGLSGVLVPRPYPYMPHSPAAAGDIAPPALS 540
Db 481 LTSLSAAERPGQMLGGLPVQMGARAGGGLSGVLVPRPYPYMPHSPAAAGDIAPPALS 540
Qy 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGAGTGIVDPNGVVLNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGAGTGIVDPNGVVLNNHVA 600
Qy 601 GATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGPPVAMGN 660
Db 601 GATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGPPVAMGN 660
Qy 661 SGGGGTTPRAVPGRWALGQTVQASDSLITGAETLNGLIQFDAAIOPGDSGGPVVNLGQ 720
Db 661 SGGGGTTPRAVPGRWALGQTVQASDSLITGAETLNGLIQFDAAIOPGDSGGPVVNLGQ 720
Qy 721 VVGMMNTAAS 729
Db 721 VVGMMNTAAS 729

RESULT 12

US-10-369-983-18
; Sequence 18, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Gunderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85b)
US-10-369-983-18

Query Match 99.8%; Score 3680; DB 15; Length 1016;
Best Local Similarity 99.9%; Pred. No. 3e-243; Indels 0; Gaps 0;
Matches 728; Conservative 0; Mismatches 1;

Qy 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAWAIGQIRSGGSPVHIGTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAWAIGQIRSGGSPVHIGTAFGLGVVD 60
Qy 61 NNGNGARVQRVVGSAAPASLSIGTGDVITAVDGAIPNSATAMADALNGHHGPDVSVTWQ 120
Db 61 NNGNGARVQRVVGSAAPASLSIGTGDVITAVDGAIPNSATAMADALNGHHGPDVSVTWQ 120
Qy 121 TKSGGTGTGNVTLAEPPAEFVDFGALPPEINSAEMVAGPAGSASIVAAQWDSVASDL 180
Db 121 TKSGGTGTGNVTLAEPPAEFVDFGALPPEINSAEMVAGPAGSASIVAAQWDSVASDL 180
Qy 181 FSAASAFQSVVWGLTVGSSWITGSSAGLWVAASPPYVAMSVTACQAEELTAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSSWITGSSAGLWVAASPPYVAMSVTACQAEELTAQVRVAAAAYE 240

Qy 241 TAYGLTVPVPIAENRAELMILITNLGQNTPAIVNNEAYGEMWAQDAAMFGYAAAT 300
Db 241 TAYGLTVPVPIAENRAELMILITNLGQNTPAIVNNEAYGEMWAQDAAMFGYAAAT 300
Qy 301 ATATATILPFEAEPMYSAGGLLEQAAAVEASDTAAANQLMNNVPOALQOLAQPTGTT 360
Db 301 ATATATILPFEAEPMYSAGGLLEQAAAVEASDTAAANQLMNNVPOALQOLAQPTGTT 360
Qy 361 PSSKLGGLWKTVPSPHRSPIINMWSMANNHSMNTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPIINMWSMANNHSMNTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
Qy 421 TAAQNGVRAMSSLSGSSLSGSSGGLGGVAANTLGRAASVGSLSVPQAWAANAQVTPAARALP 480
Db 421 TAAQNGVRAMSSLSGSSLSGSSGGLGGVAANTLGRAASVGSLSVPQAWAANAQVTPAARALP 480
Qy 481 LTSLSAAERPGQMLGGLPVQMGARAGGGLSGVLVPRPYPYMPHSPAAAGDIAPPALS 540
Db 481 LTSLSAAERPGQMLGGLPVQMGARAGGGLSGVLVPRPYPYMPHSPAAAGDIAPPALS 540
Qy 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGAGTGIVDPNGVVLNNHVA 600
Db 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVAGAGTGIVDPNGVVLNNHVA 600
Qy 601 GATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGPPVAMGN 660
Db 601 GATDINAFSVSGGQTYGVVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGPPVAMGN 660
Qy 661 SGGGGTTPRAVPGRWALGQTVQASDSLITGAETLNGLIQFDAAIOPGDSGGPVVNLGQ 720
Db 661 SGGGGTTPRAVPGRWALGQTVQASDSLITGAETLNGLIQFDAAIOPGDSGGPVVNLGQ 720
Qy 721 VVGMMNTAAS 729
Db 721 VVGMMNTAAS 729

RESULT 13

US-10-369-983-17
; Sequence 17, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Gunderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1022
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB102tm2F (MTB102FTW2, MTB72F-HTCC#1)
US-10-369-983-17

Query Match 99.8%; Score 3680; DB 15; Length 1022;
Best Local Similarity 99.9%; Pred. No. 3e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAWAIGQIRSGGSPVHIGTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAWAIGQIRSGGSPVHIGTAFGLGVVD 60
Qy 61 NNGNGARVQRVVGSAAPASLSIGTGDVITAVDGAIPNSATAMADALNGHHGPDVSVTWQ 120

Db 61 NNGNGARVQVRVGSAPAAAGISGTGDIVTAVDGPINSATAMADALNGHHPGDVSVTWQ 120
QY 121 TSGGTRTGNVTLAEGPPAEFVDFGALPPEINSAARMYAGPSASLVAAQWMDSVASDL 180
Db 121 TSGGTRTGNVTLAEGPPAEFVDFGALPPEINSAARMYAGPSASLVAAQWMDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSMIGSSAGLWVAASPYVAMSVTGAQAEHTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSMIGSSAGLWVAASPYVAMSVTGAQAEHTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVPIAENRAELMILITATNLLGNTPTAIVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVPIAENRAELMILITATNLLGNTPTAIVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGGLLEQAAAVEEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Db 301 ATATATLLPFEAPEMTSAGGLLEQAAAVEEASDTAAANQLMNNVPOALQOLAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHMTNSGVSMTNTLSSMLKGFAPAAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHMTNSGVSMTNTLSSMLKGFAPAAAARQAVQ 420
QY 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
QY 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSVLRVPPRYVMPHSPAAGDIAPPAALS 540
Db 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSVLRVPPRYVMPHSPAAGDIAPPAALS 540
QY 541 QDRPADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGAGTGIVIDPNGVLTNNHVA 600
Db 541 QDRPADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGAGTGIVIDPNGVLTNNHVA 600
QY 601 GATDINAFVSGSGQTYGVNDVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
Db 601 GATDINAFVSGSGQTYGVNDVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
QY 661 SGGQGGTPRAVPRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
Db 661 SGGQGGTPRAVPRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
QY 721 VVGMMNTAAS 729
Db 721 VVGMMNTAAS 729

RESULT 14
US-10-369-983-16
; Sequence 16, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1154
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB114F (MTB72F-ATCC#2)
US-10-369-983-16

Query Match 99.8%; Score 3680; DB 15; Length 1154;
Best Local Similarity 99.9%; Pred No. 3.5e-243;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MHHEHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSGTIVHIGTAFGLGVVD 60
Db 1 MHHEHHHTAASDNFQLSQGGGFAIPIGQAMAIAGQIRSGGSGTIVHIGTAFGLGVVD 60
QY 61 NNGNGARVQVRVGSAPAAAGISGTGDIVTAVDGPINSATAMADALNGHHPGDVSVTWQ 120
Db 61 NNGNGARVQVRVGSAPAAAGISGTGDIVTAVDGPINSATAMADALNGHHPGDVSVTWQ 120
QY 121 TSGGTRTGNVTLAEGPPAEFVDFGALPPEINSAARMYAGPSASLVAAQWMDSVASDL 180
Db 121 TSGGTRTGNVTLAEGPPAEFVDFGALPPEINSAARMYAGPSASLVAAQWMDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSMIGSSAGLWVAASPYVAMSVTGAQAEHTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGSMIGSSAGLWVAASPYVAMSVTGAQAEHTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVPIAENRAELMILITATNLLGNTPTAIVNEAEYGEWMAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVPIAENRAELMILITATNLLGNTPTAIVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGGLLEQAAAVEEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Db 301 ATATATLLPFEAPEMTSAGGLLEQAAAVEEASDTAAANQLMNNVPOALQOLAQPTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHMTNSGVSMTNTLSSMLKGFAPAAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHMTNSGVSMTNTLSSMLKGFAPAAAARQAVQ 420
QY 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPOAAANQAVTPAARALP 480
QY 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSVLRVPPRYVMPHSPAAGDIAPPAALS 540
Db 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSVLRVPPRYVMPHSPAAGDIAPPAALS 540
QY 541 QDRPADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGAGTGIVIDPNGVLTNNHVA 600
Db 541 QDRPADFPALPLDPSAMVAQVGVVNTKLYNNNAVAGAGTGIVIDPNGVLTNNHVA 600
QY 601 GATDINAFVSGSGQTYGVNDVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
Db 601 GATDINAFVSGSGQTYGVNDVGYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGN 660
QY 661 SGGQGGTPRAVPRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
Db 661 SGGQGGTPRAVPRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLGQ 720
QY 721 VVGMMNTAAS 729
Db 721 VVGMMNTAAS 729

RESULT 15
US-09-886-349A-18
; Sequence 18, Application US/09886349A
; Publication No. US20040096523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009070US
; CURRENT APPLICATION NUMBER: US/09/886,349A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: US 09/597,796
; PRIOR FILING DATE: 2000-06-20

Wed Jul 7 12:10:32 2004

; PRIOR APPLICATION NUMBER: US 60/265,737
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72FmutSA
; OTHER INFORMATION: (Ra12-TbHp-Ra35MucSA)
; US-09-886-349A-18

Query Match 99.8%; Score 3677; DB 12; Length 729;
Best Local Similarity 99.7%; Pred. No. 3.1e-243;
Matches 727; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MHHHHHTAASDNFQSLQSGGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
Db 1 MHHHHHTAASDNFQSLQSGGQGFAPIGQAMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
Qy 61 NNGNGARVQVWGVSAPASIGSTGDIVITAVDGPINSATAVADALNGHHPGDVISVTWQ 120
Db 61 NNGNGARVQVWGVSAPASIGSTGDIVITAVDGPINSATAVADALNGHHPGDVISVTWQ 120
Qy 121 TKSGGTRTGNVTLAGPPAEFVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDL 180
Db 121 TKSGGTRTGNVTLAGPPAEFVDFGALPPEINSARMYAGPGSASLVAAAQMWDSVASDL 180
Qy 181 FSAASAFQSVVWGLTVGWSWGTSSAGLWVAASPYVAVWMSVTAGCAELTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVVWGLTVGWSWGTSSAGLWVAASPYVAVWMSVTAGCAELTAAQVRVAAAAYE 240
Qy 241 TAYGLTVPPVPIAENRAELMILITNLGNTPAIAVNEAEYCEMWAQDAAMFGYAAAT 300
Db 241 TAYGLTVPPVPIAENRAELMILITNLGNTPAIAVNEAEYCEMWAQDAAMFGYAAAT 300
Qy 301 ATATATLLPPEEAPEMTSAGLLLEQAAVEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Db 301 ATATATLLPPEEAPEMTSAGLLLEQAAVEASDTAAANQLMNNVPOALQOLAQPTQGT 360
Qy 361 PSSKLGGLWKTVPSPHRSPIENMYSMANNHMTNSGVSMNTNLSMLKGFAPAPAAARQAVQ 420
Db 361 PSSKLGGLWKTVPSPHRSPIENMYSMANNHMTNSGVSMNTNLSMLKGFAPAPAAARQAVQ 420
Qy 421 TAAQNGVRAMSSLGSSLGSSGGLGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALP 480
Db 421 TAAQNGVRAMSSLGSSLGSSGGLGGVAAANLGRAASVGSLSVPOAWAANQAVTPAARALP 480
Qy 481 LTSLSAERGPQGMGLGGLPVQNGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
Db 481 LTSLSAERGPQGMGLGGLPVQNGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
Qy 541 QDRFADFPALPLDPSAMVAQVGPQVNVNITKLGYNNAVGAGTGIVIDPNGVLTNNHVIA 600
Db 541 QDRFADFPALPLDPSAMVAQVGPQVNVNITKLGYNNAVGAGTGIVIDPNGVLTNNHVIA 600
Qy 601 GATDINAFSVSGQTYGVVDVDRDQVAVLQIRGAGGLPSAAGVAVGEPVVMGN 660
Db 601 GATDINAFSVSGQTYGVVDVDRDQVAVLQIRGAGGLPSAAGVAVGEPVVMGN 660
Qy 661 SGGQGGTFRAPVGRVVALGOTVQASDSLGTGAETLNGLIQFDAAIQPGSGGPPVNVGLGQ 720
Db 661 SGGQGGTFRAPVGRVVALGOTVQASDSLGTGAETLNGLIQFDAAIQPGSGGPPVNVGLGQ 720
Qy 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:31 ; Search time 19.0527 Seconds
(without alignments)
3680.509 Million cell updates/sec

Title: US-09-597-796C-12
Perfect score: 3686
Sequence: 1 MHHEHHHTAASDNFQLSQGG.....SGGPVNVGLGVGVGMNTAAS 729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues
Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1944	52.7	391	2 B70608	probable PPE prote
2	1656.5	44.9	396	2 H70741	probable PPE prote
3	1583	42.9	393	2 C70568	probable PPE prote
4	989	26.8	355	2 F70983	probable serine pr
5	778.5	21.1	361	2 S47170	hypothetical prote
6	775.5	21.0	393	2 G70929	probable PPE prote
7	754.5	20.5	354	2 A87242	probable secreted
8	753.5	20.4	409	2 A70932	probable PPE prote
9	737	20.0	423	2 B70931	probable PPE prote
10	734	19.9	421	2 H87056	PPE-family protein
11	733.5	19.9	403	2 H70931	probable PPE prote
12	702	19.0	408	2 G70925	probable PPE prote
13	700	19.0	413	2 F70560	probable PPE prote
14	697	18.9	391	2 B70625	probable PPE prote
15	689	18.7	468	2 B70932	probable PPE prote
16	687.5	18.7	463	2 C70931	probable PPE prote
17	674	18.3	380	2 A70646	probable PPE prote
18	670.5	18.2	394	2 G70881	probable PPE prote
19	664.5	18.0	385	2 H70503	probable PPE prote
20	635	17.2	350	2 H70929	probable PPE prote
21	633.5	17.2	365	2 F70929	probable PPE prote
22	615.5	16.7	402	2 A70882	probable PPE prote
23	605.5	16.4	423	2 C70582	probable PPE prote
24	601.5	16.3	391	2 D70922	probable PPE prote
25	597.5	16.2	394	2 A70504	probable PPE prote
26	584	15.8	406	2 E70675	probable PPE prote
27	570.5	15.5	391	2 A70663	probable PPE prote
28	520.5	14.1	3300	2 D70575	probable PPE prote
29	476.5	12.9	3716	2 E70969	probable PPE prote

30	473	12.8	180	2 G70834	probable PPE prote
31	463	12.6	655	2 A70931	probable PPE prote
32	462.5	12.5	580	2 G70570	probable PPE prote
33	458	12.4	3157	2 B70969	probable PPE prote
34	452	12.3	2523	2 F70846	probable PPE prote
35	451	12.2	963	2 B70524	probable PPE prote
36	450	12.2	678	2 A70762	probable PPE prote
37	449	12.2	346	2 H70874	probable PPE prote
38	447.5	12.1	582	2 F70675	probable PPE prote
39	443	12.0	479	2 D70676	probable PPE prote
40	443	12.0	552	2 D70604	probable PPE prote
41	442	12.0	487	2 E70830	probable PPE prote
42	437.5	11.9	615	2 E70663	probable PPE prote
43	429.5	11.7	645	2 B70825	probable PPE prote
44	426.5	11.6	1053	2 F70987	probable PPE prote
45	425.5	11.5	590	2 E70946	probable PPE prote

ALIGNMENTS

RESULT 1

B70608
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: B70608
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: B70608
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-391 <COL>
A;Cross-references: GB:Z93777; GB:Al123456; NID:G3261726; PIDN:CAB07839.1; PID:G311073.1
A;Experimental source: strain H37RV
C;Genetics:
A;Gene: PPE

Query Match	52.7%	Score	1944	DB	2	Length	391
Best Local Similarity	99.7%	Pred. No.	4e-92				
Matches	390	Conservative	0	Mismatches	1	Indels	0
Gaps	0						
Qy	142	MVDFGALPPEINARMYAGPGSASLVAAACQMWDSVASDLFSAASAFQSVVWGLTVGSWIG	201				
Db	1	MVDFGALPPEINARMYAGPGSASLVAAACQMWDSVASDLFSAASAFQSVVWGLTVGSWIG	60				
Qy	202	SSAGLMVAAASPYVAMSVTAGOALTAQVRVAAAAYETAYGLTVPPVPIAENRAELMI	261				
Db	61	SSAGLMVAAASPYVAMSVTAGOALTAQVRVAAAAYETAYGLTVPPVPIAENRAELMI	120				
Qy	262	LIATNLGONTPTAINEAEYGEWMAQDAANEGYAAATATATATALLPEEPAPMTSAGG	321				
Db	121	LIATNLGONTPTAINEAEYGEWMAQDAANEGYAAATATATATALLPEEPAPMTSAGG	180				
Qy	322	LLRQAAVFEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN	381				
Db	181	LLRQAAVFEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN	240				
Qy	382	MVSMANNHSMNTNSGVSMNTNTLSSMLKGFAPAAARQAVOTAQNGVRAMSSLGSSG	441				
Db	241	MVSMANNHSMNTNSGVSMNTNTLSSMLKGFAPAAARQAVOTAQNGVRAMSSLGSSG	300				
Qy	442	LGCGVAAVFEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN	501				
Db	301	LGCGVAAVFEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHSPISN	360				
Qy	502	GQMGAPAGGLGCVLRVPPRPPYMPHSPAAG	532				

```

Db 361 GQMGARAGGSLGVLVRPPRPVMPHSPAAG 391
RESULT 2
H70741
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: H70741
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: H70741
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-396 <COL>
A: Cross-references: GB:275555; GB:AL123456; NID: g2261608; PID: e250360;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: PPE

Query Match 44.9%; Score 1656.5; DB 2; Length 396;
Best Local Similarity 85.1%; Pred. No. 1.8e-77;
Matches 338; Conservative 18; Mismatches 34; Indels 7; Gaps 3;

QY 142 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVGLTVGSMIG 201
Db 1 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVGLTVGSMIG 60

QY 202 SSAGLVAAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 261
Db 61 SSAGLVAAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 120

QY 262 LIATNLGQNTPAIVNEAYEGWQAADAAAFGYAAATATATATLLPPEEAPMTSAGG 321
Db 121 LIATNLGQNTPAIVNEAYEGWQAADAAAFGYAAATATATATLLPPEEAPMTSAGG 180

QY 322 LLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTGCTTPSSKLGGLWKTVPSPHSPISN 381
Db 181 LLEQAAVEEADITAAANQLMNNVPOALQOLAQPTKSIWPFQDLSBLWKAISPHLSPLSN 240

QY 382 MVSNNHNSMTNSGVSMNTTSSMLKGFAPAARQAVOTAAQNGVRMSS----LGSSL 437
Db 382 MVSNNHNSMTNSGVSMNTTSSMLKGFAPAARQAVOTAAQNGVRMSS----LGSSL 437

QY 438 GSSGLGGVAANLGRAASVGLSVPOAWAANAQAVTPAARALPLTSLTSAERGPQOMLG 497
Db 300 GSSGLGGVAANLGRAASVGLSVPOAWAANAQAVTPAARALPLTSLTSAERGPQOMLG 359

QY 498 GLPVGMGARAG--GGLSGVLVRPPRPVMPHSPAAG 532
Db 360 GLPLGQLTNSGGGFGVSNALRMPPRAYVPRVPAAG 396

RESULT 3
C70568
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: C70568
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: C70568
A: Status: preliminary; nucleic acid sequence not shown; translation not shown

Query Match 26.8%; Score 989; DB 2; Length 355;
Best Local Similarity 98.0%; Pred. No. 1.5e-43;
Matches 196; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 530 AAGDIAPALSDQDFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVGAGTGVIDPN 589
Db 28 APAQAAAPALSDQDFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVGAGTGVIDPN 87

QY 590 GVLVTNNHVITAGATDINAFSVGSCQTGVGVYDRFDQVAVLQLRGAGLPSNAIGGV 649

```

```

A: Molecule type: DNA
A: Residues: 1-393 <COL>
A: Cross-references: GB:295390; GB:AL123456; NID: g2261766; PID: CAB08702.1; PID: e316074;
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: PPE

Query Match 42.9%; Score 1583; DB 2; Length 393;
Best Local Similarity 81.5%; Pred. No. 9.7e-74;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 142 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVGLTVGSMIG 201
Db 1 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVGLTVGSMIG 60

QY 202 SSAGLVAAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 261
Db 61 SSAGLVAAASPVVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIAENRAELMT 120

QY 262 LIATNLGQNTPAIVNEAYEGWQAADAAAFGYAAATATATATLLPPEEAPMTSAGG 321
Db 121 LIATNLGQNTPAIVNEAYEGWQAADAAAFGYAAATATATATALLPPEEAPMTSAGG 180

QY 322 LLEQAAVEEASDTAAANQLMNNVPOALQOLAQPTGCTTPSSKLGGLWKTVPSPHSPISN 381
Db 181 LLEQAAVEEADITAAANQLMNNVPOALQOLAQPTGCTTPSSKLGGLWKTVPSPHSPISN 240

QY 382 MVSNNHNSMTNSGVSMNTTSSMLKGFAPAARQAVOTAAQNGVRMSS----LGSSL 437
Db 241 VSSIANNHNSMTNSGVSMNTTSSMLKGLAPAAA-QAVTPAENGVRMSSLSGSLGSSL 299

QY 438 GSSGLGGVAANLGRAASVGLSVPOAWAANAQAVTPAARALPLTSLTSAERGPQOMLG 497
Db 300 GSSGLGGVAANLGRAASVGLSVPPAWAANAQAVTPAARALPLTSLTSAAQTPAGHMLG 359

QY 498 GLPVGMGARAGGGLSGVLVRPPRPVMPHSPAAG 532
Db 360 GLPLIGH-SVNAGSGINNALRVPARAYAIPTPAAG 393

```

RESULT 4

```

F70983
probable serine proteinase pepA - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002
C:Accession: F70983
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: F70983
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-355 <COL>
A: Cross-references: GB:296071; GB:AL123456; NID: g3242254; PID: CAB09453.1; PID: g2181967
A: Experimental source: strain H37RV
C: Genetics:
A: Gene: pepA
C: Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; trypsin

```

```

Query Match 26.8%; Score 989; DB 2; Length 355;
Best Local Similarity 98.0%; Pred. No. 1.5e-43;
Matches 196; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 530 AAGDIAPALSDQDFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVGAGTGVIDPN 589
Db 28 APAQAAAPALSDQDFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVGAGTGVIDPN 87

QY 590 GVLVTNNHVITAGATDINAFSVGSCQTGVGVYDRFDQVAVLQLRGAGLPSNAIGGV 649

```

Db 88 GYVLTNNHVIAGATDINAFSVSGQTYGVVDVYDRTQDVAVLQRLGAGGLPSAAGGV 147
QY 650 AYGEFVWAMNSGGGGTTPRAVGRVWALQGVQASDLSLTGAETLNGLIQFDAAIQGD 709
Db 148 AYGEFVWAMNSGGGGTTPRAVGRVWALQGVQASDLSLTGAETLNGLIQFDAAIQGD 207
QY 710 SGGPVVNGLGQVGVNMNTAAS 729
Db 208 SGGPVVNGLGQVGVNMNTAAS 227

RESULT 5
S47170
hypothetical protein 34k - Mycobacterium paratuberculosis
C:Species: Mycobacterium paratuberculosis
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Dec-2002
C:Accession: S47170
R: Cameron, R.M.; Stevenson, K.; Inglis, N.F.; Klausen, J.; Sharp, J.M.
submitted to the EMBL Data Library, June 1993
A:Description: Isolation and characterisation of a 34kDa protein of Mycobacterium paratu
A:Reference number: S47170
A:Accession: S47170
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-361 <CM>
A:Cross-references: EMBL:723092; NID:G505550; PIDN:CAA80638.1; PID:G505551
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

Query Match 21.1%; Score 778.5; DB 2; Length 361;
Best Local Similarity 66.0%; Pred. No. 8.1e-33;
Matches 155; Conservative 28; Mismatches 37; Indels 15; Gaps 2;

QY 495 MLGGLPVGMQARAGGLSGVLVRPVRPVMPHSPAAAGDIAPPALSDRFPALPDDP 554
Db 15 LVGLVTVWGLGLSGVG-----LAPASA--APSGALDRFADRLPIADP 59
QY 555 SAMVAGVGPQVNVNITKLYNNNAVAGAGTGIVIDPVGVLNTHNVIAGATDINAFVSGGQ 614
Db 60 SAMVAGVGPQVNVNITKLYNNNAVAGAGTGIVIDPVGVLNTHNVIAGATDINAFVSGGQ 119
QY 615 TYGVVDVYDRTQDVAVLQRLGAGGLPSAAIGGVAVGVPVWAMNSGGGGTTPRAVGR 674
Db 120 TVAVDVYDRTQDVAVLQRLGAGGLPSAAIGGVAVGVPVWAMNSGGGGTTPRAVGR 179
QY 675 VVALGQTVQASDLSLTGAETLNGLIQFDAAIQGDVSGGVPVWAMNSGGGGTTPRAVGR 729
Db 180 VVALGQTVQASDLSLTGAETLNGLIQFDAAIQGDVSGGVPVWAMNSGGGGTTPRAVGR 234

RESULT 6
G70929
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70929
R: Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, C.; Churcher, C.; Harris, D.; Gordon, S.
Connor, R.; Davies, K.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70929
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-393 <COL>
A:Cross-references: GB:AL020201; GB:AL123456; NID:G3250699; PIDN:CAAL17711.1; PID:el25460
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 21.0%; Score 775.5; DB 2; Length 393;
Best Local Similarity 43.4%; Pred. No. 1.3e-32;

Matches 178; Conservative 65; Mismatches 130; Indels 37; Gaps 10;
QY 143 VDFGALPPEINSEARMYAGPGSASLVAAQWDSVADLFSAAQFQVWGLTVGWSIGS 202
Db 1 MDFGALPPEVNSVRYMAGPGSAPVWAAASAWNGLAELSSAATGYETVITQLSSEGWLGP 60
QY 203 SAGLVMAAASPYVWMSVTAGQAEELTAAQVVRVAAAAYETAYGLTVPPVTAENRAELMIL 262
Db 61 ASAAAEAVAPYVWMSAAAQAEQATQAAAAAFEAFAATVPPPLIAANRASLMQL 120
QY 263 IATNLLGONTPALJAVNEAEYGENWAOAAAAAFYAAATATATATLTPFEEAPEMTSAGGL 322
Db 121 ISTNVFGONTSAIAAAEAYGENWAOQDSAAAYAYAGSSASASA-VTPFSTPQIANPTAQ 179
QY 323 LEQAAAVEBASDTAAA--NOLMNVPOALQOLAAQPGQTTPSSKLGGLWK----- 370
Db 180 GTQAAAATAGTAQSTLTETMITGPNALQSLTSLPLQSS-NGPLSLWMLQILFETFPNPT 238
QY 371 -----TVSPHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVCTAA 423
Db 239 SISALLTDLPQYASFFVNTGLPYFSGMGNNFIQSAKTL-GLIGSNAPAAV-----AAA 292
QY 424 QNGVRAMSSLSGSSGLGGVVAANLGRAASVGSLSVPOAWAA--ANQAVTPAARALPLT 482
Db 293 GDAAKGLPGLGMLG---GCPVAAAGLGNAAASVGLSVPPVWMSGPIPGSVTPGAAPLVS 348
QY 483 SLTSAARERPGQMLGGLPVQMGARAGGSLSVLRVPPRYVWPHSPAAAG 532
Db 349 TVSAAPAEAPFGSLGGLPL----AGAGGAGAGP-RYGFRTVMARPPFAG 393

RESULT 7
A87242
probable secreted serine proteinase [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002
C:Accession: A87242
R: Cole, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, F.R.; Hor
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
cam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Skelton, J.; Squares, R.; Sq
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: A87242
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-354 <STO>
A:Cross-references: GB:AL450380; NID:gl3093863; PIDN:CAC32191.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML2659
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; try

Query Match 20.5%; Score 754.5; DB 2; Length 354;
Best Local Similarity 73.3%; Pred. No. 1.3e-31;
Matches 151; Conservative 25; Mismatches 25; Indels 5; Gaps 2;

QY 524 VWPSPAGDIAPPALSDRFPALPDPSPAMVAGVGPQVNVNITKLYNNNAVAGAGT 583
Db 27 VWPSP--ATSPGPTLALDRFSNPPLNPAWVA---PQVNVNITKLYNNNAVAGAGT 81
QY 584 IVIDPVGVLNTHNVIAGATDINAFVSGGQTYGVVDVYDRTQDVAVLQRLGAGGLPSA 643
Db 82 IVIDSSGGVLTNNHVISGATDISAFDVGNGKTYGVVDVYDRTQDVAVLQRLGAGGLPSA 141
QY 644 AIGGVAVGVPVWAMNSGGGGTTPRAVGRVWALQGVQASDLSLTGAETLNGLIQFDA 703
Db 142 VIGDVAIGEPVAGNTGGGGLPSVLPGRVVALNQTQVQASEPLTGAQETLSGLIQVDA 201
QY 704 AIQFGDSGGPVVNGLGQVGVNMNTAAS 729
Db 202 PIKFGDSGGPVVNSRGQVGVNMNTAAT 227

RESULT 8

A70932
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: A70932
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: A70932
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-409 <COL>
A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAAL17729.1; PID:e125461
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 20.4%; Score 753.5; DB 2; Length 409;
Best Local Similarity 41.8%; Pred. No. 1.8e-31;
Matches 184; Conservative 52; Mismatches 139; Indels 65; Gaps 12;

QY 143 VDFGALPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFOSVVMGLTVGWSIGS 202
DB 1 MDFGALPEINSGRMYAGPGSGPLLAHAAWDAALAEALYSAASYSSTIEGLTVAFWMG 60
QY 203 SAGLMVAAASPYVAMSVTAQOAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMIL 262
DB 61 SSIITMAAAVAPYVAMISVTAGQAEQAGAKIAAGVYETAFAATVPPVIAENRALLMSL 120
QY 263 IATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLLPFEAPMTSAGGL 322
DB 121 VATNIFGQNTPAIATBAHVAEMWAQDAAMYGAGSSATA-SQLAPFEPPTNPSAT 179
QY 323 LEQAAVEEASDTAA-----NLMNNVPQALQQLAQTPTTSSKLGGLWKTVSPH 375
DB 180 AAQSAVVAQAAGAAASDDITAOQLSLLPSTLQSLA--TTATATSASAG--WDTVL-- 233
QY 376 RSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAA-----ARQAVQTAQ 424
DB 234 -----QSITITLANITGYSIIGLAIPEGWMLTFGQILGLAQNAPGVAA 279
QY 425 NGVR-----AMSSL-----GSSILGS-SGLGGVAAANLGRAASVGLSVFQWAAANQAVTPA 475
DB 280 LGPKAAAGALSPLAPLGGVIGDITPLGGATGCIARAIIVGSLSVFQGWAAEAPVWRAV 339
QY 476 ARALPLTSLTSA-AERPGQMLGGLPVGMGARGAGGL-----SGVLRVPPRYVMPHSP 529
DB 340 ASVLPGFGGAAPALAAEAPGALFEMALSSLAGRALAGTAVRSAGAARV----- 388
QY 530 AAGDIAPPALSDQRFADFFPA 549
DB 389 AGGSVTEDEVASTTIIIVIPA 408

RESULT 9

B70931
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: B70931
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: B70931

A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-423 <COL>
A:Cross-references: GB:AL022021; GB:AL123456; NID:g3250699; PIDN:CAAL17722.1; PID:e125461
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 20.0%; Score 737; DB 2; Length 423;
Best Local Similarity 41.8%; Pred. No. 1.3e-30;
Matches 183; Conservative 50; Mismatches 125; Indels 80; Gaps 12;

QY 143 VDFGALPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFOSVVMGLTVGWSIGS 202
DB 1 MDFGALPEINSGRMYTGPFGPMLAATAWDGLAVELHATAAGVASELSALT-GAWSGP 59
QY 203 SAGLMVAAASPYVAMSVTAQOAEELTAQVRAAAAYETAYGLTVPPVIAENRAELMIL 262
DB 60 SSTSMASAAAPYVAMSVATAVHAELAGQAARLAIAAYEAAFAATVPPVIAENRAELMVL 119
QY 263 IATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLLPFEAPMTSAGGL 322
DB 120 IATNIFGQNTPAIMTAEQYMEWMAQDAAMYGAGSSATA-SMTATFEPPTNNGQL 178
QY 323 LEQAAVEEASDTAAAN-----OLMNNVPQALQQLAQTPTT-----OQTTP-----S 362
DB 179 GAQSSAVAQTAATAAGNLQSAFFQALLSAVPRALQGLALETASQASATPQWTDLGNLS 238
QY 363 SKLGLWKTVSPHRSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAARQAVQTA 422
DB 239 TFLGG--AVTGPYTP-----GVLPSPGVPLGLIQSVL-----V 271
QY 423 AQNGVRAAMSSILGS-----SLGSSGLGG--VAANLGRAASVGLS 460
DB 272 TONGQGVSAALLGKIGKFTGALAPLAEFALHTPLGSEGLGGSVSAGIGRAGLVGKLS 331
QY 461 VPOWAAANQAVTPAARALPLTSLTS---AERPGQMLGGLPVGMGARGAGGLSGVLR 517
DB 332 VFGQMTVAAPIPSPAAALQATRLAARAPIAATDGAGALLGMLSLAGRAAAGSTG--- 388
QY 518 VPPRYVMPHSPAAAGDIA 535
DB 389 ---HPIGSAAAPAVGAA 403

RESULT 10

H87056
PPE-family protein [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: H87056
R: Cole, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holroyd, R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd, R.; M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; St
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: H87056
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-421 <STO>
A:Cross-references: GB:AL405380; NID:g13093150; PIDN:CAC31563.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML1182

Query Match 19.9%; Score 734; DB 2; Length 421;
Best Local Similarity 41.0%; Pred. No. 1.8e-30;
Matches 175; Conservative 54; Mismatches 156; Indels 42; Gaps 9;

QY 142 MVDGALPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFOSVVMGLTVGWSIG 201

Db 1 MFDAALSPETNSTRMVYLGSSPILTAATAAAWVVLAKELTAAAGLQSAVEAL-LTTFEG 59
QY 202 SSANGLMVAASPVVAMSVTAQAEHTAAQVVAATAAAYETAYGLTVPPPIAENAEELMI 261
Db 60 ESAALAEVTPYKELMTQNAASAEUTATQLTVAANAYETATMTVPPPLMVFVNRQAACL 119
QY 262 LIAINLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPMTSAGG 321
Db 120 LIMSIFGQNSTAIAEKEAEYETMTIQDAAAVTSYQASVLEAVGATKAPTAPPLGVNEVG 179
QY 322 L-----LQAAAVEASDTAAANQLMNN-----VPOALQ-----LAQP 355
Db 180 LAQEVVEVVEVVEVVEVVEVVEVVEVVEVVEVVEVVEVVEVVEVVEVVEVVEVVEV 239
QY 356 TQGTTPSSKLG--LWKTVPSPHSPISNMVMANNHMTNSGVSM-TNTLS-SMLKGFAPA 413
Db 240 QTAVPDSSAAAPQLMGGAQAHLSPINDTLNHNHAGVANAGLSLVNGWGSAMKSLAPT 299
QY 414 ARQAVQTAONGVRAMSLGSSLG-----SSGLGGVAAANLGRAASVGLSVPOAWAAN 469
Db 300 TTK-----AAESAFKAMGSAVSTGRGLGSSSGSHVTAGLGRAASIGSLRVPQTWTAS 354
QY 470 QAVTPAARALPLTSLTSAARERPGOML-GGLPVGQM--GARAGGGLSGVLVPPRPVYM 525
Db 355 QPVTATRALSPARVAVATESAPLLGGGLPWMPVPGSGGTGGVNTALRLQPRAFVM 414
QY 526 PHSPAAG 532
Db 415 PRNPAAG 421

RESULT 11
H70931
probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: H70931
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: H70931
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-403 <COL>
A: Cross-references: GB:AL022021; GB:AL123456; NID: G3250699; PIDN: CAAL7728.1; PID: el25461
A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: PPE

Query Match 19.9%; Score 733.5; DB 2; Length 403;
Best Local Similarity 42.9%; Pred. No. 1.9e-30;
Matches 181; Conservative 62; Mismatches 124; Indels 55; Gaps 13;

QY 143 VDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGWSIGS 202
Db 5 LDFATLPPPEINSARMYAGSAPMLAASAAHGLSALRASLSYSSVLSTUTGEEWGP 64
QY 203 SAGLWAAASPYVAMSVTAQAEHTAAQVVAATAAAYETAYGLTVPPPIAENAEELMI 262
Db 65 ASASMTAAAPYVAMSVTAQAEHTAAQVVAATAAAYETAYGLTVPPPIAENAEELMI 124
QY 263 IATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 322
Db 125 IATNVLGONAPATAAATAEAQVAAEWMSQDAMMYAGASAAAT-QLTFPTEPQTNNASGL 183
QY 323 LEQAAA-----VEBASDTAAANQLMNN-----VP--QALQ--LAQPTQGTTPSSKL 365
Db 184 AQAASAAIAHAGSAGNAQTTLSQLIAIPSVLQGLSSSTAAATFASPGSLGIVCGSS 243

QY 364 KLGLWKTVSPHRSPISNMVMANNHMTNSGVSM-TNTLS-SMLKGFAPA-AAARQA 418
Db 244 WLDKMLALDPN-----SNFNTIASSGLFLPSNTIAPFLGLGGVAAAADAAGDV 293
QY 419 VQTAONGVRAMSLGSSLGSS-GLGGVAAANLGRAASVGLSVPOAWAANQAVTFAAR 477
Db 294 LGEATSGGLG--GALVAPLGSAGLGTGAAGLGNAAATVGLTSLVPPSTAAAPLASHLGS 351
QY 478 AL-----PLTSLTSAARERPGOML-GGLPVGQM-GARAGGGLSGVLVPP-----RPYVMPHSPA 530
Db 352 ALGGTPMVAPPVAAAG-----MPCMPFGTGGGQFG-----RAVPOYGFEPNFAVPPA 401
QY 531 AG 532
Db 402 AG 403

RESULT 12
G70925
probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70925
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A: Reference number: A70500; MUID: 98295987; PMID: 9634230
A: Accession: G70925
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-408 <COL>
A: Cross-references: GB:Z74024; GB:AL123456; NID: G3250700; PIDN: CAA98377.1; PID: el301025;
A: Experimental source: strain H37Rv
C: Genetics:
A: Gene: PPE

Query Match 19.0%; Score 702; DB 2; Length 408;
Best Local Similarity 42.3%; Pred. No. 7.4e-29;
Matches 180; Conservative 48; Mismatches 144; Indels 54; Gaps 13;

QY 143 VDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGWSIGS 202
Db 1 MDFGVLPPEINSARMYAGPGSGPMMAAAWDSLAELGLAAGGYRLAISLTGAYWAGP 60
QY 203 SAGLWAAASPYVAMSVTAQAEHTAAQVVAATAAAYETAYGLTVPPPIAENAEELMI 262
Db 61 AAASVAAVTPYVAMSVTAQAEHTAAQVVAATAAAYETAYGLTVPPPIAENAEELMI 120
QY 263 IATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPPEEAPMTSAGG 322
Db 121 VATNFFGONTPAIAATEAQVAAEWMAQDAAMMYAGASAAAT-ELTPTTAAPVTISPAAL 179
QY 323 LEQAAA-----VEBASDTAAANQLMNN-----VP--QALQ--LAQPTQGTTPSSKL 365
Db 180 AQAATAVTSVTPPLATTAAYVPLLQQLSSSTSLIPWYSALQWLAEALLGLTDPNMTIV 239
QY 366 -----GGLKTVSPHRSPISNMVMANNHMTNSGVSM-TNTLS-SMLKGFAPA 413
Db 240 RLLGISYFDEGL-----LQFEASLAQQAIPGTPGGAG--DSGSSVLDSWGPTIFA 287
QY 414 AARQAVQTAONGVRAMSVSLGS-----SLGSSGLGGVAAANLGRAASVGLSVPOAWAAN 469
Db 288 GPRASPSVAGGAGVGGVOTPOPYWYWDRESIGSVSAALGKSSAGSLSPVPPDWAARA 347
QY 470 QAVTPAARALP-----LTSITSAARPGOML-GGLPVGQVGAARAGGGLSGVLVPPRPVYM 526
Db 348 RWANPAWELPQDDVTALRGTAENA---LLRGFFMASAGQSTGGGF--VHKYGFRLAVNQ 402
QY 527 HSPAAG 532

Db 403 RPPFAG 408

RESULT 13

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: F70560

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID: 98295987; PMID: 9634230

A: Accession: F70560

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-413 <COL>

A: Cross-references: GB:295436; GB:AL123456; NID:G3261770; PIDN: CAB08826.1; PID: e316565;

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: PPE

Query Match 19.0%; Score 700; DB 2; Length 413;

Best Local Similarity 39.1%; Pred. No. 9.5e-29;

Matches 168; Conservative 60; Mismatches 146; Indels 56; Gaps 7;

QY 142 MVDGALPPEINSARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 201

DB 1 MLDPAQLPPEINSALMAGPGSGPMLAAAEWELAELOTTASTDYDALTGLADGPWQG 60

QY 202 SSAGLMAAASPYVAMSVTAGOAELETAQVRVAAAYETAYGLTVPPPIAENRAELMI 261

DB 61 SSSAASVAAATPQVWLRSTAGQAEQAGSQVAAASAYEAFFATVPVPPPIAENRALLMA 120

QY 262 LIATNLGONTPTAIVNEAEYEGEMWQADAAAGYAAATATATATATLLPPEAPEMTSAGG 321

DB 121 LLATNFLGONTPTAATAETQAEVNEAWQADAAAGYAAATATATATATLLPPEAPEMTSAGG 179

QY 322 LLEQAAVEEASDSTAANQ-----LNNVPOALQLOAQPTQGTTPSSKLGGLKMTVSPHRSPI 381

DB 180 LASQAASVQGVASGAANAQALTDIPKAL-----FGLSGIFTNEPFWLTDLIGK 226

QY 382 MYSMANNHMTNSGVSKTNTLSMLKGFAPAA---ARQAVQT----- 421

DB 227 ALGUTGHTWSDGSLTVGGVDFVQGTGSAELDASVADTFGKWSPARLMVTQFKD 286

QY 422 -----AAQNGVRAMSLSGLSSGLGGGVAANTLGRAASVGSLSVPOA 464

DB 287 YFGLAHDLPKWSGKAGAEAKALPAAPVAPSPAGL-SGVAGAVQAAASVGLKVPVAV 345

QY 465 WAAANQAVTPAARALPTLSAARQFGQMLGGLPVQGVQAGARAGGSLSVL--RVPPRP 522

DB 346 WTATTTPAASPAVLAAASGLGAAAEAGSTHAFGGMPL--MGSGAGRAFNFAAPRYGFKP 403

QY 523 YVMPHSPAAG 532

DB 404 TVIAQPPAGG 413

RESULT 14

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: B70625

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID: 98295987; PMID: 9634230

A: Accession: B70625

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-468 <COL>

A: Cross-references: GB:AL022021; GB:AL123456; NID:G3250699; PIDN: CAA17730.1; PID: e1254620;

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: PPE

Query Match 18.7%; Score 689; DB 2; Length 468;

Best Local Similarity 38.4%; Pred. No. 4e-28;

Matches 174; Conservative 66; Mismatches 147; Indels 66; Gaps 11;

QY 143 VDFGALPPEINSARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 202

DB 1 MDFGLQPEITSGEYMLPGAGPMLAAAVAVDGLAAELQSWAASVYASIVEGMASESMLGP 60

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID: 98295987; PMID: 9634230

A: Accession: B70932

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-468 <COL>

A: Cross-references: GB:AL123456; NID:G3250699; PIDN: CAA17730.1; PID: e1254620;

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: PPE

Query Match 18.7%; Score 689; DB 2; Length 468;

Best Local Similarity 38.4%; Pred. No. 4e-28;

Matches 174; Conservative 66; Mismatches 147; Indels 66; Gaps 11;

QY 143 VDFGALPPEINSARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 202

DB 1 MDFGLQPEITSGEYMLPGAGPMLAAAVAVDGLAAELQSWAASVYASIVEGMASESMLGP 60

RESULT 15

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: B70932

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squires, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID: 98295987; PMID: 9634230

A: Accession: B70932

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-468 <COL>

A: Cross-references: GB:AL022021; GB:AL123456; NID:G3250699; PIDN: CAA17730.1; PID: e1254620;

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: PPE

Query Match 18.7%; Score 689; DB 2; Length 468;

Best Local Similarity 38.4%; Pred. No. 4e-28;

Matches 174; Conservative 66; Mismatches 147; Indels 66; Gaps 11;

QY 143 VDFGALPPEINSARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 202

DB 1 MDFGLQPEITSGEYMLPGAGPMLAAAVAVDGLAAELQSWAASVYASIVEGMASESMLGP 60

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and activities. It emphasizes that this is crucial for ensuring transparency and accountability in the organization's operations. The text also mentions that proper record-keeping is essential for identifying trends and making informed decisions.

2. The second part of the document outlines the various methods used to collect and analyze data. It describes how data is gathered from different sources and how it is processed to extract meaningful insights. The text highlights the importance of using reliable and valid data sources to ensure the accuracy of the findings.

3. The third part of the document focuses on the results of the data analysis. It presents the findings in a clear and concise manner, using tables and graphs to illustrate the data. The text also discusses the implications of the findings and how they can be used to improve the organization's performance.

4. The fourth part of the document provides a summary of the key findings and conclusions. It reiterates the importance of maintaining accurate records and using reliable data sources. The text also offers recommendations for future research and actions to be taken by the organization.

5. The final part of the document is a conclusion that summarizes the overall findings and conclusions. It emphasizes the importance of continuous monitoring and evaluation to ensure the organization remains effective and efficient. The text also expresses confidence in the organization's ability to achieve its goals and objectives.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:42:21 ; Search time 11.8689 Seconds

(without alignments)
3198.204 Million cell updates/sec

Title: US-09-597-796C-12

Perfect score: 3686

Sequence: 1 MHHRHHHTAASDNFOLSGG.....SGGPWNLGQVGMNTAAS 729

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 segs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1656.5	44.9	396	1 YD61 MYCTU	Q11031 mycobacteri
2	702	19.0	408	1 YS92 MYCTU	Q10813 mycobacteri
3	687.5	18.7	463	1 Y102 MYCTU	O53951 mycobacteri
4	450	12.4	678	1 YF48 MYCTU	Q10778 mycobacteri
5	445	12.1	487	1 Y442 MYCTU	P42611 mycobacteri
6	424.5	11.5	443	1 Y878 MYCTU	Q10540 mycobacteri
7	363	9.8	408	1 SRA MYCLE	Q07297 mycobacteri
8	338	9.2	434	1 YU18 MYCTU	P31500 mycobacteri
9	332.5	9.0	435	1 YU21 MYCTU	O53268 mycobacteri
10	331	9.0	463	1 Y056 MYCTU	Q10892 mycobacteri
11	262.5	7.1	458	1 YVPA BACSU	Q9R911 bacillus su
12	233.5	6.3	178	1 YV29 MYCTU	O06246 mycobacteri
13	228	6.2	458	1 HRA2 HUMAN	O43464 homo sapien
14	225.5	6.1	448	1 DEGB ARATH	Q91u10 arabidopsis
15	225.5	6.1	864	1 ELS RAT	Q93172 rattus norv
16	225	6.1	355	1 DEGS ECOLI	P31137 escherichia
17	225	6.1	488	1 DEGP CHLPH	Q926t0 chlamydia p
18	222	6.0	437	1 DEGI ARATH	O22609 arabidopsis
19	222	6.0	455	1 DEQ2 ECOLI	P39099 escherichia
20	221.5	6.0	458	1 HRA2 MOUSE	O91j55 mus musculu
21	221	6.0	449	1 HTPA BACSU	Q43358 bacillus su
22	220.5	6.0	860	1 ELS MOUSE	P54320 mus musculu
23	219.5	6.0	400	1 YVXA BACSU	P39668 bacillus su
24	219.5	6.0	513	1 DEGP BRUSU	Q45597 brucella su
25	218.5	5.9	478	1 DEGP BUCAI	P57322 bucherna ap
26	217.5	5.9	176	1 YV25 MYCTU	Q50703 mycobacteri
27	217	5.9	466	1 HTOA HAEIN	P45129 haemophilus
28	215.5	5.8	2432	1 Y43R TRV6	P8305 chilo iride
29	215	5.8	413	1 HTEA LACHE	Q924h7 lactobacill
30	215	5.8	508	1 DEGP RICCN	Q92jal rickettsia
31	213	5.7	730	1 ELS HUMAN	P15502 homo sapien
32	211.5	5.7	513	1 DEGP BRUME	Q8Yg32 brucella me
33	211	5.7	475	1 DEGP SALTY	P26982 salmonella

ALIGNMENTS

RESULT 1

ID	YD61 MYCTU	STANDARD;	PRT;	396 AA.
AC	Q11031:			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Hypothetical PPE-family protein Rv1361c/MT1406.			
GN	Rv1361C OR MT1406 OR MTCY02B10.25C.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=H37RV;			
RC	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornisby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.,			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544(1998).			
[2]				
SEQUENCE FROM N.A.				
RC	STRAIN=CDC 1551 / Oshkosh;			
RC	MEDLINE=22206494; PubMed=12218036;			
RX	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,			
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.,			
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains";			
RL	J. Bacteriol. 184:5479-5490(2002).			
CC	-!- SIMILARITY: Belongs to the mycobacterial PPE family.			
CC				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC				
DR	EMBL; Z75555; CAA99966.1; -			
DR	EMBL; AE007013; AK45669.1; -			
DR	PIR; H70741; H70741.			
DR	TIGR; MT1406; -			
DR	TuberculList; Rv1361c; -			
DR	InterPro; IPR000030; Microbac_PPE.			

Q9fec4 chlamydomon
P04985 bos taurus
P09376 escherichia
P18584 chlamydia t
P35828 caulobacter
O85291 bucherna ap
Q9p197 chlamydia m
Q52894 rhizobium m
O05942 rickettsia
P15921 rickettsia
P54925 bartonella
P12021 sus scrofa

DR Pfam; PF00823; PPE; 1.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 158 159 TA -> AT (IN REF. 2).
 SQ SEQUENCE 396 AA; 40015 MW; 6AFAED07B5F68D0 CRC64;

Query Match 44.98; Score 1656.5; DB 1; Length 396;
 Best Local Similarity 85.18; Pred. No. 2.4e-74;
 Matches 338; Conservative 18; Mismatches 34; Indels 7; Gaps 3;

QY 142 MVDFFGALPPPEINSARMYAGPGSASLVAAQMDVSDLFSAASAFQSVVWGLTVGSGWG 201
 DB 1 MVDFFGALPPPEINSARMYAGPGSASLVAAQMDVSDLFSAASAFQSVVWGLTVGSGWG 60
 QY 202 SSAGLMVAASPYVAMSVTGAQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 261
 DB 61 SSAGLMVAASPYVAMSVTGAQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 120
 QY 262 LIATNLGQNTPAIAVNEAEYGMWQADAAAFGYAAATATATATATLLPFEAPEMTSAGG 321
 DB 121 LIATNLGQNTPAIAVNEAEYGMWQADAAAFGYAAATATATATALLPFEAPLIINPGG 180
 QY 322 LLGQAAVEBASDTAAANQMLNNVPQALQLOAQTOGTTPSSKLGGLWKTVPSPHRSPIIN 381
 DB 181 LLEQAVAEBAIDTAANQMLNNVPQALQLOAQTKSIWPFDDQLSELWKAISPHLSPLSN 240
 QY 382 MVSMANNHMTNSGVSMNTTLLSMLKGFAPAPAAQAVOTAAQNGVRAMSS----LGSSL 437
 DB 241 IVSMNNHVMNTNSGVSMNTTLLSMLKGFAPAPAAQAVETAAQNGVQVAMSSLLGSSL 299
 QY 438 GSSGLGGVGAANTGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTGAABRGPGQMUG 497
 DB 300 GSSGLGAGVAAANTGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTGAATAPGHMUG 359
 QY 498 GLPVGOMGARAG--GGLSGLVRLVPPPPYVMPHSPAAAG 532
 DB 360 GLPLGQJTNSSGGFGVGSNALRMPPRAYVMPRVPAAAG 396

RESULT 2
 YS92 MYCTU STANDARD; PRT; 408 AA.
 AC Q10813;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein RV2892C/MT2959/MB2916c.
 GN RV2892C OR MT2959 OR MTCY274.23C OR MB2916C.
 OS Mycobacterium tuberculosis, and
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773, 1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
 RX MEDLINE=9829598; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Horsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL J. Bacteriol. 184:5479-5490(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Gordon S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Akkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis."
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL; Z74024; CAA98377.1; -.
 CC EMBL; AE007119; AAK47285.1; -.
 CC EMBL; BX248344; CAD96603.1; -.
 CC FIR; G70925; G70925.
 CC TIGR; MT2959; -.
 CC TubercuList; RV2892c; -.
 CC InterPro; IPR00030; Microbac_PPE.
 CC Pfam; PF00823; PPE; 1.
 CC Hypothetical protein; Transmembrane; Complete proteome.
 KW TRANSMEM 56 76 POTENTIAL.
 FT SEQUENCE 408 AA; 41469 MW; 3E3D1P20D7827199 CRC64;
 SQ

Query Match 19.0%; Score 702; DB 1; Length 408;
 Best Local Similarity 42.3%; Pred. No. 9.7e-28;
 Matches 180; Conservative 48; Mismatches 144; Indels 54; Gaps 13;

QY 143 VDFGALPPEINSARMYAGPGSASLVAAQMDVSDLFSAASAFQSVVWGLTVGSGWG 202
 DB 1 MDFGVLPEINSGRMYAGPGSGPMWAAAAAASLAAELGLAAGYRLAISLTCAYWAGP 60
 QY 203 SAGLMVAASPYVAMSVTGAQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 262
 DB 61 AAASWVAATPYVAWLSATAGAQEQAGQAAAAAAAYELAFAMTVPVPPVVAANRALVAL 120
 QY 263 IATNLGQNTPAIAVNEAEYGMWQADAAAFGYAAATATATATLLPFEAPEMTSAGG 322
 DB 121 VATNFFGQNTPAIAATEAQVAEMWQADAAAYAGSAAIAT-ELTPTTAAPVTTSPAAL 179
 QY 323 LEQAAA-----VEEASDTAAANQMLNN-----VP--QALQQ-LAQPTQGTTFSSKL--- 365
 DB 180 AQQAATVSSVTPPLATTAAPVQLQLLSLIPWYSAALQQLAENLLGLTDPNRMTIV 239
 QY 366 -----GGLWKTVPSPHRSPIINMYSMANNHMTNSGVSMNTTLLSMLKGFAP---A 413
 DB 240 RLLGTSYFDEGL-----LQFEASLAQQAIPGTPGGAG--DSGSSVLDGSGPTIFA 287
 QY 414 AARGAVOTAQNGVRAMSSLSGS-----SLGSSGLGGVAAANLGRAASVGSLSVFOQMAAAN 469
 DB 288 GPRASPSVAGGAGVGGVQTPQPIYWNALDRESIGSVAALGKSSAGSLSVPPDWARA 347
 QY 470 QAVTPAARALP---LTSLTGAABRGPGQMUGLGLPVGOMGARAGGLSVLRVPPRVPMP 526
 DB 348 RWANPAAWRLPGDDVTALRGTAENA---LLRGFFPMASAGQSTGGGF--VHKYGFRLAVMQ 402
 QY 527 HSPAAG 532
 DB 403 RPPFAG 408

```
RESULT 3
YI02_MYCTU
AC 053951;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rv1802/MT1851/Wb1830.
GN Rv1802 OR MT1851 OR MTW049.24 OR Wb1830.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jégels K., Krogh A., McLean J., Moule S., Murphy J.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.Dovis; STRAIN=AF122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin J., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AL022021; CAAL1723.1; --
CC EMBL; A5007044; AAK4123.1; --
CC EMBL; EX248340; CAD94533.1; --
CC PIR; C70931; C70931.
CC TIGR; MT1851; --
CC Tuberculist; Rv1802; --
CC InterPro: IPR000030; Microbac_PPE.
CC Pfam: PF00823; PPE; 1.
CC Hypothetical protein; Complete proteome.
CC CONFLICT 401 401
CC
```

```
SQ SEQUENCE 463 AA; 46021 MW; EE64828BF09FA551 CRC64;
Query March 18.7%; Score 687.5; DB 1; Length 463;
Best Local Similarity 42.8%; Pred. No. 5.7e-27;
Matches 174; Conservative 48; Mismatches 136; Indels 49; Gaps 11;

QY 143 VDFGALPPEINSMYAGPGSASLVAAQWDSVADLPSAASAFQSVVWGLTVGWSIGS 202
DB 1 MDEGVLPPPEINSGRMVAGPGSGPMLAAAAAANDGLATELQSTADYGSVISVLT-GVMSGQ 59
QY 203 SAGIWAASAPYVAMVMSVTAGQAELETAQVRVAAAYETAYGLTVPPVIAENRAELMIL 262
DB 60 SSGTMAAAPYVAMVMSVTAALAREAAQAASAAAAAYEAFAATVPPVVAANRAELAVL 119
QY 263 IATNLLGQNTPAIVNEAEYGENWQDAAMFCYAAAAATATATATLLPFEAPEMTSAGGL 322
DB 120 AATNIFQNTGATAAAAEARYAEMWQDAAMYGAGSSSVAT-QVTFFAAPPTTTNAAGL 178
QY 323 LEQAAVEEASDTAAANQLMNVTPQALQLOAQTOGTPTSSKLGGLWKTYS--PHRSPI- 379
DB 179 ATQGVAVQAQVAGSAGN-ARSLVSEVLEFLA--TAGTNNYKTKVASLMNAVTVGFYASSVY 235
QY 380 -----SNMVSMMNNHMSMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQAQGVRA 429
DB 236 NSMLGLGFABSKVLPANDIVISTIFGMVQFOKFFNPVTFNF-----DLIP 282
QY 430 MSSLGSSLG-----SSGLGG---GVAANLGRAASVGSLSVPOAWAANONAVTPAARALPL 481
DB 283 KSALGAGLGLRSASISGLSGSTAPASAGASQAGSVGMSVPPPSWAAATFAIRTVAAVFSS 342
QY 482 TSLTS--AERGPQML-----GGLPVGQNGARAGGGLSGVLRV 518
DB 343 TGLQAVPAAAISSGSLLSQVALASVAGGALGGAARATGGLGGGRV 389

RESULT 4
YF48_MYCTU
ID_YF48_MYCTU STANDARD; PRT; 678 AA.
AC Q10778;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUN-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rv1548c/MT1599.
GN Rv1548c OR MT1599 OR MTCY48.17.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jégels K., Krogh A., McLean J., Moule S., Murphy J.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT
```



```

[3]
RN  SEQUENCE OF 160-374 FROM N.A.
RP  STRAIN=Isolate 50410;
RC  Paki A.H., Dale J.W.;
RL  Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.
CC  -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC  -!- CAUTION: In strain Oshkosh the gene for this protein is
CC  interrupted in position 307 by an 186110 element.
CC  -!- CAUTION: was originally (Ref.3) thought to be a dihydrofolate
CC  reductase.
CC  -!- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
CC  in positions 294; 337 and 355.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; AL021287; CA416103.1; --
CC  EMBL; AE007129; AAK47427.1; ALT SEQ.
CC  EMBL; AE007129; AAK47430.1; ALT SEQ.
CC  EMBL; X59271; CA41961.1; ALT_FRAME.
CC  PIR; E70857; E70857.
CC  TIGR; MT3098; --
CC  TIGR; MT3101; --
CC  TubercuList; RV3018C; --
CC  InterPro; IPR00030; Microbac_PPE.
CC  Pfam; PF00823; PPE; 1.
CC  Hypothetical protein; Complete proteome.
KW  SEQUENCE 434 AA; 43029 MW; 41D673C4BD389DD6 CRC64;
SQ
Query Match          9.2%; Score 338; DB 1; Length 434;
Best Local Similarity 27.2%; Pred. No. 6.1e-10;
Matches 123; Conservative 54; Mismatches 211; Indels 64; Gaps 12;

QY  147 ALPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIGSSAGL 206
DB  8 ASPPEVHSGALLSAGPGPSLQAAAGWGSALSAAVAQELSVVAAVAGAGVWQGPSAEL 67

QY  207 MYAASPYVAMSVMTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMILIATN 266
DB  68 FVAATVPVYVAMVQASADSAAAGHEAAGVVCALAEMLTLPLEANHLTHAVLVATN 127

QY  267 LIGQNTPAIVNEARYGEMWAQDAAMFGYAAATATATATLLPFEEAPEMTSAGLLLEQA 326
DB  128 PFGINTIPALNEADYVRMWQAATVMSAYEAVVGAALVATPHTGPAFVIVKPG----- 181

QY  327 AAVEASDTAAAN-----QLMNVNPOALQQAQTOGTPSSKLGGLWKTVSFH 375
DB  182 --ANEASNAVAATAITPPFPHHIVQFLEETFAAYDOYLSALLSELPA--VAMVWFQLFVD 237

QY  376 --RSPISNNVSMANNHSMNTSGVSWNTLSSMLKGFAPAAARQAQVOTAAQNGVRAMSS 432
DB  238 ILGFNIIGIITLASNAQLLTFEAINASVAVGLLYAIA-GVIDIVVEWVIGNLFGVVPL 296

QY  433 LGSSL-----GSSGLGG--GVAA--NLGRAASVGSLSVPQAAANAAQVTPAARA 478
DB  297 LGGPLLALGAAVAVVPGVAGLAGVAGLAALPAPVGAAGAPAAALVGVVAPVSGVWSPQAR- 355

QY  479 LPLTSLTSAERGGPGQMLGSLPGVQNGARAGG--GLSGVLVPPPPVYVPHSPHAPAGDIAP 536
DB  356 -----LVSAVEFAPASTSVSVLSADRGALGFVGTAG-----KESVGPQAG 397

QY  537 PALSQDRFADFALPLDPSAMVAQVGPQVNI 568
DB  398 LTVLADEFDGGAPVPMPLGWSW----GPDLVGV 425

RESULT 9
YU21_MYCTU

```

```

YU21_MYCTU STANDARD; PRT; 435 AA.
AC 053268; C53269;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV3021C/RV3022C/MT3106.
GN RV3021C/RV3022C OR MT3106 OR MTV012.35C/MTV012.36C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H3RV;
RC MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RC MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Paterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Emoliaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RA "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 82.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL021287; CA416106.1; ALT_FRAME.
CC EMBL; AL021287; CA416107.1; ALT_FRAME.
CC EMBL; AE007129; AAK47435.1; --
CC TIGR; MT3106; --
CC TubercuList; RV3021C; --
CC InterPro; IPR00030; Microbac_PPE.
CC Pfam; PF00823; PPE; 1.
CC Hypothetical protein; Complete proteome.
CC CONFLICT 299 299 G->A (IN REF. 2).
CC CONFLICT 317 320 LAGV->VTGL (IN REF. 2).
CC CONFLICT 326 326 L->V (IN REF. 2).
CC SEQUENCE 435 AA; 42876 MW; 3B157643EAA8484A CRC64;
SQ
Query Match          9.0%; Score 332.5; DB 1; Length 435;
Best Local Similarity 25.3%; Pred. No. 1.1e-09;
Matches 123; Conservative 59; Mismatches 172; Indels 133; Gaps 14;

QY  147 ALPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIGSSAGL 206
DB  8 ASPPEVHSGALLSAGPGPSLQAAAGWGSALSAAVAQELSVVAAVAGAGVWQGPSAEL 67

QY  207 MYAASPYVAMSVMTAGQAEITAAQVRAAAAYETAYGLTVPPVIAENRAELMILIATN 266

```

Db 68 FVAAYVFAVLQASADSAAGAEHEAAAGYVCALEMPITLPELAANHLTHAVLVATN 127
 QY 267 LLGQNTPAIAVNEAEYGEWAQDAANAFGYAA 298
 Db 128 FFGINTPIALNEADYVVMVQATVMSAYEAVVGAALVATPHGTGPAPVIVKPGANEASN 187
 QY 299 ATATATATLPPFE-----APEMTSAGGLLEQAAVEEASDAAANOLMNNVPOALQ 350
 Db 188 AVAAATITPPFGLAKLEFARQAFTVEGELIMKSAEAWAVFVLTGLNVFEP----- 243
 QY 351 QLAQTOGTTPSSKIGLWKTVPSPHRPISNNVSM-----ANNH 389
 Db 244 -----MLV-----LTCMIDMPFATVGFALGVFVLPVLPLEPAVILE 278
 QY 390 MSMTNSGVSMNTLSS-----MLKGFAPAAARQAVCTAAQNGVRAMSLGSSLSGSL 442
 Db 279 LAILSIGWIIISNFGAIPVLGGELLGMAAAVVPVAGLGVAGLAALPAVGAAGAP-- 336
 QY 443 GGGVAANLGRAASV-GSLSVPOAAANQAVTPAARALPLTSLTSAERGPQMLGLPV 501
 Db 337 -----AALVGVAPVCGVGSPQARLVLS--AVEPAPASTSVSLV--ASDRGAGAL--GF-V 385
 QY 502 GQVGARAGGSLGVLRPPPYVMPHSPAGDIAPALSDRPAFPALPLDPSAMVAQV 561
 Db 386 GTAKESVGQAPGL-----TVIADEFGDGPVFMPLPGSW----- 419
 QY 562 GPQVNI 568
 Db 420 GPDVGV 426

RESULT 10

Y096 MYCTU STANDARD; PRT; 463 AA.
 AC Q10852;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PFE-family protein Rv0096/MT0105.
 GN Rv0096 OR MT0105 OR MTCY251.15.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekoa F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Ormrod L.A., Ermolaeva M., Salzberg S.L.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL J. Bacteriol. 184:5479-5490 (2002).
 RL J. Bacteriol. 184:5479-5490 (2002).
 CC -|- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -|- SIMILARITY: Belongs to the mycobacterial pPE family.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; Z74410; CAA98932.1; -.
 DR EMBL; AE006922; AAK44327.1; -.
 DR TIGR; H70750; H70750.
 DR TIGR; MT0105; -.
 DR Tuberculist; Rv0096; -.
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF0823; PPE; 1.
 DR Hypothetical protein; Transmembrane; Complete proteome.
 KW TRANSMEM 3 23 POTENTIAL.
 FT TRANSMEM 88 108 POTENTIAL.
 FT TRANSMEM 112 132 POTENTIAL.
 FT TRANSMEM 216 236 POTENTIAL.
 FT TRANSMEM 245 265 POTENTIAL.
 FT TRANSMEM 276 296 POTENTIAL.
 FT TRANSMEM 323 343 POTENTIAL.
 FT TRANSMEM 419 439 POTENTIAL.
 SQ SEQUENCE 463 AA; 46894 MW; 42D9D66A033D0DD8 CRC64;
 Query Match 9.0%; Score 331; DB 1; Length 463;
 Best Local Similarity 25.2%; Pred. No. 1.4e-09;
 Matches 137; Conservative
 QY 147 ALPEINSAEMYAGPGSALVAAAQMWDSVASLFSAAAFQSVVMGLTVGSMIGSSAGL 206
 Db 2 AIPEVHSGLLSAGCGPGLLVAAQWQELSDQVALACAEGLQGLGEVQASSWQTAATQ 61
 QY 207 MVAASPYVAMSVTACQAELETAQVRAAAVETAVGLTVPPVPAENPAELMILATN 266
 Db 62 YVAAHGPYLAWEQTAINSATVAAQHVAAAACVCSLAAMPFAELAAHAIHGVLIATN 121
 QY 267 LLGQNTPAIAVNEAEYGEWAQDAANAFGYAAATATATATLLPFEAPEMTSAGGLLEQA 326
 Db 122 FFGINTVPIALNEADYVVMVQATVMSAYEAVVGAALVATPHGTGPAPVIVKPGANEASN 176
 QY 327 AAVEEASDT-----AAANQLMNNVPOALQALQPTQ-----GTPPSKL----- 365
 Db 177 -----DAADTRLDVLSISGQIRDI---LDFIANPKYKYLEFFQGFSPAVTVVVALVAL 239
 QY 366 ---GGLW-----KTVSPHRGPISNMVSMANNHSMNTNSGVSMNTLSSMLK-----GF 410
 Db 230 QLYDFLWYPPYASVGLLLLPFFTP-----TLALTALSAHLNLNLPAGL 275
 QY 411 APAARQAVQTAQNGVRAMSLGSSLSGSGGVAAANLGRAASVGSLSVPOAANAQ 470
 Db 276 LPFAA-----ALGPGDQWGANLA-----V 294
 QY 471 AVTPAARALPLTSLTSAERGPQMLGLGVQMGARAGGSLGVLRVPPVPMHPSA 530
 Db 295 AVTPATAAVP-----GGSPPTNPAPAPSSNSVGSASNAPOI---SYA 335
 QY 531 AGDIAPPALSDRFA--DFFPALDPSAMVAQVGPQVNVINTKLGYNNNAVAGTGTVIDP 588
 Db 336 VPLGAPGVSSGPKAGTKSPDTAATLTATAGAARFGLAHRKRKRSSEGVG-----IRGY 390
 QY 589 NGVVLNNHVIAGATDINAEVSGSGTYGVVGVYDETQDVAVLQLRGAGLPSAIGGG 648
 Db 391 RDEFLDATTATVDAATDVPANAAG--SQAGTGLF-----AGTAPTTSS---G 433
 QY 649 VAVG 652
 Db 434 AAAG 437

RESULT 11
YVTA_BACSU STANDARD; PRT; 458 AA.
ID Q9R311; Q95021; Q35039;
AC 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable serine protease yvta (EC 3.4.21.-).
GN YVTA OR YVTB OR BSU33000.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1] _SEQUENCE FROM N.A.
RP STRAIN=168;
RC STRAIN=168;
RX MEDLINE=20158875; PubMed=10692364;
RA Noone D., Howell A., Devine K.M.;
RT "Expression of ykda, encoding a Bacillus subtilis homologue of HtrA,
RT is heat shock inducible and negatively autoregulated.";
RL J. Bacteriol. 182:1592-1599 (2000).
RN [2]
RP _SEQUENCE FROM N.A.
RX MEDLINE=9804033; PubMed=9394377;
RA Kunst F., Ogasawara N., Mosser I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertsch M.G., Bessieres P., Bolotin A., Borchert S.,
RA Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Guisepi G., Guy B.J., Haga K., Haiesch J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lindholm S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan B., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takenaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vanier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weizenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.P., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis.";
RL Nature 390:249-256 (1997).
RN [4]
RP TRANSCRIPTIONAL REGULATION.
RC STRAIN=168;
RX MEDLINE=20576168; PubMed=11133960;
RA Noone D., Howell A., Collier R., Devine K.M.;
RT "ykda and yvta, HtrA-like serine proteases in Bacillus subtilis,
RT engage in negative autoregulation and reciprocal cross-regulation of
RT ykda and yvta gene expression.";
RL J. Bacteriol. 183:654-663 (2001).
CC -1- FUNCTION: May be involved in processing, maturation, or secretion
CC of extracellular enzymes.

CC -1- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -1- INDUCTION: Induced by heat shock during exponential growth and by
CC heterologous amylases at the transition phase of the growth cycle.
CC Negatively regulates its own expression.
CC -1- MISCELLANEOUS: Inactivation results in compensating overexpression
CC of htrA, especially during stress conditions.
CC -1- SIMILARITY: Belongs to peptidase family S2C.
CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -1- CAUTION: Ref. 2 and Ref. 3 sequences differ from that shown due to
CC frameshifts in positions 87 and 246 that produce two separate
CC ORFs.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF188296; AAF03153.1; -
CC EMBL; Z93941; CAB07968.1; ALT_FRAME.
CC EMBL; Z93941; CAB07969.1; ALT_FRAME.
CC EMBL; Z93120; CAB15290.1; ALT_FRAME.
CC SUBTILIS; SGL4155; Yvta.
CC InterPro; IPR009003; Cys_ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_S1C.
CC Pfam; PF00595; PDZ_1.
CC Pfam; PF00089; tryptsin; 1.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SMO0228; PDZ; 1.
CC PROSITE; PS50106; PDZ; 1.
CC Hydrolase; Protease; Serine protease; Heat shock; Transmembrane;
CC Complete proteome.
CC DOMAIN 1 71 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 72 92 POTENTIAL.
CC DOMAIN 93 458 EXTRACELLULAR (POTENTIAL).
CC DOMAIN 356 440 PDZ.
CC ACT_SITE 187 187 CHARGE RELAY SYSTEM (POTENTIAL).
CC ACT_SITE 217 217 CHARGE RELAY SYSTEM (POTENTIAL).
CC ACT_SITE 298 298 CHARGE RELAY SYSTEM (POTENTIAL).
CC SEQUENCE 458 AA; 48717 MW; 77551045A865A5CD CRC64;
Query Match 7.1%; Score 262.5; DB 1; Length 458;
Best Local Similarity 29.1%; Pred. No. 3.le-06;
Matches 85; Conservative 52; Mismatches 116; Indels 39; Gaps 14;
QY 459 LSVQAAANQAVTPAARALPLTSLTSAERPGQMLGGI-PVGOMGARAGGSL-GVL 516
DB 38 LDAPVSYEGRQ---ETASALEMEKQETAVKKEKRAAWLSPI--LGGIGGGLMGI- 91
QY 517 RVPRPYVPHSPFAAGDIAPPALSDQRFADFPALPLDPSA---MVAQVGPQVNI-NTK 571
DB 92 ---APVLPDQNOATETA-SANKQVQSDNFTTAPITNASIADWEDLEPTIVGISNIQ 146
QY 572 LGYNAV-----GAGTGIVT---DPNGVLTNNHVIAGATDINAFVSGQTY 616
DB 147 TSQNTFTGCGSSSESGTGSGVIFKQSDKAYIINHHVVEGANKLTV-TLYNGETE 205
QY 617 GVDVVGDRTOAVLQLRGAGGLPSAAIGGG--VAVGEPVVMNGSGGQGTTPRAVPR 674
DB 206 TAKLVGSDTTDLAVLEISGNVKKVASFGDSSQLRTGKVIAGNPLGQFSGTQTQGI 265
QY 675 VVALGQTVQASDLTGAETLNGLIQDAIQPDGSGGPPVNGLGQVVGWNT 726
DB 266 ISGLNRTIDV-DTTQGTVM--NVLTQDAAINPGSGPLINASSGQVIGINS 314
RESULT 12
Y129_MYCTU
ID Y129_MYCTU STANDARD; PRT; 178 AA.

RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[6]
RP SEQUENCE OF 134-458, INTERACTION WITH BIRC4, AND MUTAGENESIS OF
ALA-134.
RX MEDLINE=21468395; PubMed=11583623;
RA Suzuki Y., Imai Y., Nakayama H., Takahashi K., Takio K., Takahashi R.;
"A serine protease, HtrA2, is released from the mitochondria and
interacts with XIAP, inducing cell death.";
Mol. Cell 8:613-621 (2001).
[7]
RN CHARACTERIZATION.
RX MEDLINE=20334437; PubMed=10873535;
RA Savopoulos J.W., Carter P.S., Turconi S., Pettman G.R., Karan E.H.,
Gray C.W., Ward R.V., Jenkins O., Creasy C.L.;
"Expression, purification, and functional analysis of the human serine
protease HtrA2.";
Mol. Cell 8:613-621 (2001).
[8]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 134-458.
RX MEDLINE=22028567; PubMed=11967569;
RA Li W., Sriivasas S.M., Chai J., Li P., Wu J.W., Zhang Z.,
Alnemri E.S., Shi Y.;
"Structural insights into the pro-apoptotic function of mitochondrial
serine protease HtrA2/Omi.";
Nat. Struct. Biol. 9:436-441 (2002).
CC -!- FUNCTION: Serine protease that shows proteolytic activity against
a nonspecific substrate beta-casein. Promotes or induces cell
death either by direct binding to and inhibition of BIRC proteins
(also called inhibitor of apoptosis proteins, IAPs), leading to an
increase in caspase activity, or by a BIRC inhibition-independent,
caspase-independent and serine protease activity-dependent
mechanism. Isoform 2 seems to be proteolytically inactive.
CC -!- SUBUNIT: Interacts with Mxi2. The mature protein, but not the
precursor, binds to BIRC2, BIRC3 and BIRC4/XIAP.
CC -!- SUBCELLULAR LOCATION: Mitochondrial, predominantly present in the
intermembrane space. Released into the cytosol following apoptotic
stimuli, such as UV treatment, and stimulation of mitochondria
with caspase-8 truncated BID/tBID.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1; Synonyms=13B;
CC IsoId=O43464-1; Sequence=Displayed;
CC Name=2; Synonyms=D-Omi;
CC IsoId=O43464-2; Sequence=VSP_005359, VSP_005361;
CC Name=3; Synonyms=p7;
CC IsoId=O43464-3; Sequence=VSP_005360, VSP_005361;
CC Name=4; Synonyms=p4;
CC IsoId=O43464-4; Sequence=VSP_005362;
CC -!- TISSUE SPECIFICITY: Isoform 1 is ubiquitous; isoform 2 is
expressed predominantly in the kidney, colon and thyroid.
CC -!- DOMAIN: The PDZ domain mediates interaction with Mxi2.
CC -!- PTM: Autophosphorylationally activated.
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF020760; AAB94569.2; -;
DR EMBL; AF141305; AAF66596.1; -;
DR EMBL; AF141306; AAF66597.1; -;
DR EMBL; AF141307; AAF66598.1; -;
DR EMBL; AF149311; AAG13126.1; -;
DR EMBL; AC065444; -; NOT_ANNOTATED_CDS.
DR EMBL; BC000096; AAB00096.1; -;
DR PDB; 1LCY; 29-MAY-02.

DR MEROPS; S01.278; -;
DR Genew; HGNC:14348; PRSS25.
DR MM; 606441; -;
DR GO; GO:0005789; C:endoplasmic reticulum membrane; TAS.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0003754; F:serine-type endopeptidase activity; TAS.
DR GO; GO:0004252; F:proteolysis and peptidolysis; TAS.
DR GO; GO:0006508; P:proteolysis and stress; TAS.
DR GO; GO:0006950; P:response to stress; TAS.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ_Ser_trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00839; trypsin; 1.
DR PRINTS; PR00834; PROTEASIN2C.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00202; Tryp_SPC; 1.
DR PROSITE; PS0106; PDZ; 1.
DR Hydrolase; Protease; Serine protease; Zymogen; Mitochondrion;
KW Transit peptide; Transmembrane; Apoptosis; Alternative splicing;
KW 3D-structure.
KW TRANSIT 1 31 MITOCHONDRION.
FT PROPEP 32 133
FT CHAIN 134 458 SERINE PROTEASE HTRA2.
FT TRANSMEM 105 125 POTENTIAL.
FT DOMAIN 166 342 SERINE PROTEASE.
FT DOMAIN 364 445 PDZ.
FT SITE 134 137 IAP-BINDING MOTIF.
FT ACT_SITE 198 198 CHARGE RELAY SYSTEM.
FT ACT_SITE 228 228 CHARGE RELAY SYSTEM.
FT ACT_SITE 306 306 CHARGE RELAY SYSTEM.
FT VASAPLIC 238 302 Missing (in isoform 2).
FT VASAPLIC 313 313 Missing (in isoform 3).
FT VASAPLIC 313 313 L -> LARELGAVSLQ (in isoform 3).
FT VASAPLIC 372 403 /FTid=VSP_005360.
FT VASAPLIC 314 458 /FTid=VSP_005361.
FT VASAPLIC 314 458 DGEVIGYVMVLTSPSLAEQLRPFSPDQVQGLIHK
FT VASAPLIC 314 458 VILGSPAHRAGLRPGDVLAIQGMQNAEDVAVETQSO
FT VASAPLIC 314 458 LAVQIRRGRTLLVTPTEVE -> VSETSLFRLPAPGQ
FT VASAPLIC 314 458 CGKRPFLIOGCLVPLSLSLAISQYTPRSPQLLVLEF
FT VASAPLIC 314 458 CPHLEFLV (in isoform 4).
FT VASAPLIC 314 458 /FTid=VSP_005362.
FT MUTAGEN 134 134 A-M: LOSS OF INTERACTION WITH BIRC4;
FT MUTAGEN 134 134 LOSS OF INHIBITION OF BIRC4 ACTIVITY.
FT MUTAGEN 134 134 A-M: LOSS OF INTERACTION WITH BIRC4 AND
FT MUTAGEN 306 306 CF INHIBITION OF BIRC4 ACTIVITY.
FT SEQUENCE 458 AA; 48840 MW; CEA955A7D0DD8C0D CRC64;
S-A: LOSS OF PROTEASE ACTIVITY.
Query Match 6.2%; Score 228; DB 1; Length 458;
Best Local Similarity 28.8%; Pred. No. 0.00015;
Matches 104; Conservative 42; Mismatches 129; Indels 86; Gaps 17;
QY 411 APAARQAVCTAAGVQVRAMSSSLGSLGGVAAVLG---RAASVGSLSVPOA--- 464
DB 3 APRAGR-----GAGSLRAWALG-----GIRGRRPRITPDRLALLTSGTSPRARVT 51
QY 465 -----WAAANQAV-----TPAARA-----LPLTSLTAAERPGQMLGGLPVGQ 503
DB 52 YGTFSLWRLSVGVTEPRACLTSGTPGPRAGLTAVTPTDTRTREAENSGTRSAWLAVA- 110
QY 504 MGARA-----GGGLSGVLVPPRPVMPHSPRAAGDIAPALSODRF---ADFPALPLD 553
DB 111 LGAGGAVLLLLWGGG-----RGPPAVLAAPVSP-----FPASPRQYNFIAD----- 152
QY 554 PSAMVAQVGPQVNI-----NTKLGYNNVAGAGTGVIDPENGVLVTKNNHVIAGATDINAF 608
DB 153 ---VVEKTAPAVVYIILDRHPLGREGVPSNGSGFVVAADGLIVTNAHVVAADRVRV- 208
QY 609 SVSGGQTYGVVDVGVDRDQDVAVLQLRGAGGLPSAAIG---GGVAVGEPVAVMGSGGGG 666

```
DB 209 RLLSGDTYEAUVTVADVPVADVIATLRIQTKFPLTLPGRSADVRQGFVVAM-----G 261
QY 667 TPRAVPGRVVA-LGQTQVQASDLTGAEETLNGLIQFDAAIQPGDSGGVGVNGLGVVGMN 725
DB 262 SPFALQNTITSGIVSSAQRARDLGLTQVNTVEYIQTDAADDFGNSGGPVLNLDGIVGVN 321
QY 726 T 726
DB 322 T 322

RESULT 14
DEG8_ARATH STANDARD; PRT; 448 AA.
AC Q9LUI0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protease Do-like 8, chloroplast precursor (EC 3.4.21.-).
GN DEG8 OR A5G39830 OR K13H13.1.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RN Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE OF 91-113 AND 345-361.
RC STRAIN=cv. Columbia;
RA Kieselbach T., Bystedt M., Schroeder W.P.;
RN Submitted (OCT-2000) to Swiss-Prot.
CC -!- FUNCTION: Probable serine protease.
CC -!- SUBCELLULAR LOCATION: Chloroplast; within the thylakoid lumen
(CC (Probable)).
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; AB024023; BAA98101.1;
DR MEROPS; S01.279;
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00634; PROTEASES2C.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS0106; PDZ; FALSE NEG.
KW Hydrolase; Serine protease; Transit peptide; Chloroplast; Thylakoid.
FT TRANSIT 1 ?
FT TRANSIT 1 ?
FT CHAIN 91 448 PROTEASE DO-LIKE 8.
FT DOMAIN 152 333 SERINE PROTEASE.
FT DOMAIN 336 433 PDZ.
FT ACT_SITE 171 171 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 214 214 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 292 292 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 448 AA; 47492 MW; A986FC1387670AFF CRG64;
```

```
Query Match 6.1%; Score 225.5; DB 1; Length 448;
Best Local Similarity 28.1%; Pred. No. 0.0002;
Matches 87; Conservative 41; Mismatches 103; Indels 79; Gaps 13;

QY 456 VGSLSVPAWAA-----ANQAVTFAARALPLTSTSAERPGQMLGGLPVGMGARAGG 510
DB 41 VSSVKVTDQWKSNLHELAVKSPSTTRILLTSLFNLCLFNPSTLYLSALALGD----- 93
QY 511 GLSGVLRVPPRPYVPHSPAAGDIAPALSDREADFPALPDPS-AMVAQVGPQ----V 565
DB 94 -----PSVATVEDVSPTV-----FPAGLFFTEGRIVOLFETKNTYSV 130
QY 566 VNI-----NTKLGVNNV-----GAGTGIVIDPENGVLVLTNNHVIAGATDIN----- 606
DB 131 VNIFFVTLRPLQKMTGVVEIPEGNGSGVWMDGQGYIVTNYHVIGNALSRNPSPGDVVGRV 190
QY 607 --AFSVGSGQTYVDVWGYDRTQDVAVLQKAGAGLPSAAI-----GGGVAVGEPVVA 657
DB 191 NILASDGVQKNEFKLVGADRAKDLAVLKV---DAPETLLKPIKVGQSNLKVQOCLA 246
QY 658 MNSGGQGGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNG 717
DB 247 IGNPFGPDHT--LTGVVISGLNRDI---FSQTV--TIGGGIQTDAAINPNSGGPLDS 299
QY 718 LGQVVGWNTA 727
DB 300 KGNLIGINTA 309

RESULT 15
ELN_RAT STANDARD; PRT; 864 AA.
AC Q99372;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Elastin precursor (Tropoelastin) (Fragment).
GN ELN.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=91104868; PubMed=1702999;
RA Pierce R.A., Deak S.B., Stollie C.A., Boyd C.D.;
RT "Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.";
RL Biochemistry 29:9677-9683(1990).
[2]
RP SEQUENCE OF 781-864 FROM N.A.
RX MEDLINE=88330868; PubMed=2971041;
RA Deak S.B., Pierce R.A., Belsky S.A., Riley D.J., Boyd C.D.;
RT "Rat tropoelastin is synthesized from a 3.5-kilobase mRNA.";
RL J. Biol. Chem. 263:13504-13507(1988).
[3]
RP SEQUENCE OF 264-533 AND 558-864 FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=92241859; PubMed=1572637;
RA Pierce R.A., Alatawi A., Deak S.B., Boyd C.D.;
RT "Elements of the rat tropoelastin gene associated with alternative
splicing.";
RL Genomics 12:651-658(1992).
CC -!- FUNCTION: Major structural protein of tissues such as aorta and
nuchal ligament, which must expand rapidly and recover completely.
CC -!- SUBUNIT: The polymeric elastin chains are cross-linked together
into an extensible 3D network.
CC -!- SUBCELLULAR LOCATION: Extracellular matrix of elastic fibers.
CC -!- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=8;
Comment=Experimental confirmation may be lacking for some
isoforms;
Name=1;
IsoId=Q99372-1; Sequence=Displayed;
```

[illegible]

Search completed: June 30, 2004, 16:49:46
Job time : 12.8689 secs

Db	422	SPAAAAKAAKAAK--YGARGVGIPYGVGAGFPYGVGAGAGL-----	466
QY	221	TAGQBELTAAQVRVAAAAAYETAAGLTVPVPPVIAENRAELMILIATNLLQONTPAIVNVEA	280
Db	466	--GASQAAAA--AAAAAKAATGAG-----GAGTLGGLVPG-----	497
QY	281	EYGEWMAQDAAMFGYAAATATATATATLLPEEAPEMTSAGLL---EQAAAVEREASDTAA	337
Db	498	-----AVPGALPGAVPCALPGAVPGALP-GAVPGVPGTGGVPGAGTAAAAAAXAAKAA	550
QY	338	ANQLMNNVPQALQOLAQTGGTTPSS-KLGGLWKTVSPHRSPISNMVSMANHHMSTMNSG	396
Db	551	A-----KAGQTLGFLGPGVGVPGGVGGVGLPGGVGP-----GVTGTG	587
QY	397	VSM-TNTLSSMLKGF-APAAAARAVOTAQNGVRAMSSLCSSLSGSSGLGGV-----AA	448
Db	588	TGPGTGLVPGLDGGAGTTPAAKSAKAAKAAKQYAAAGLGAAGVPLGVGAGVPGFCAGAG	647
QY	449	NLGRAASV---GSLVSQVQAAANAQVTPAARALPITSLTSAERGPQWGLGLPVQMG	505
Db	648	QFGAGAGVPGFCAGAVPGSLAAKAAKYAAGSL-----GGPGLGAGPPGGLGGPG	697
QY	506	ARAG-GGLSGVLYRPPRPVYVMPHSPAAAGDIAPALSCDRFADFPALPLDPSAMVAQVGPQ	564
Db	698	GFGPGGLGGV-----PGVAGG--APAAAAAKAA-----AKAAQYG--	733
QY	565	VVNINTKLYNNAVGAGTGIVDPNGVVLTNNHVIAGTDINAFSVSGQTGVGVVGVYD	624
Db	734	-----LGGAGGLGAG-GL-----GAGGLGAGGLGAG-----GLGAGSL--	765
QY	625	RTQDVAVLQLRGAGL-----PSAAI-----GGGVAVG--	652
Db	766	-----GAGGVTPGAVGLGGSVPAKAAKAAKYAGLGGVLGAPFFCGGVAAARP	814
QY	653	-----EPVYAMNGSG---QCGTTPRAVGRVVALG	679
Db	815	GFGLSPIYFGGGAGLGVGGKPPKPYGCGALGALG	848

1. The first part of the document is a letter from the President of the United States to the Congress, dated January 3, 1863. It is a very important document, as it contains the President's message to Congress regarding the state of the Union and the progress of the war. The letter is written in a formal, official style, and is signed by Abraham Lincoln.

2. The second part of the document is a report from the Secretary of War, dated January 10, 1863. It provides a detailed account of the military operations and the state of the army during the previous year. The report is written in a more detailed and technical style, and is signed by Edwin M. Stanton.

3. The third part of the document is a report from the Secretary of the Treasury, dated January 15, 1863. It provides a detailed account of the financial operations and the state of the treasury during the previous year. The report is written in a more detailed and technical style, and is signed by Alexander C. Gibson.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:01 ; Search time 53.0977 Seconds
(without alignments)
4331.879 Million cell updates/sec

Title: US-09-597-796C-12
Perfect score: 3686
Sequence: 1 MHHHHTAASDNFOLSOGG.....SGGPVNGLGQVGVNNTAAS 729

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues 1017041
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL.25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_invertebrate:*
5: sp_mammal:*
6: sp_mhc:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1944	52.7	391	16	O05298 mycobacteri
2	1920.5	52.1	390	16	Q7U0E9 mycobacteri
3	1779.5	48.3	396	16	Q7U071 mycobacteri
4	1648	44.7	393	16	Q8VIZ3 mycobacteri
5	1583	42.9	393	16	O06341 mycobacteri
6	1583	42.9	393	16	Q7TWFS mycobacteri
7	989	26.8	355	16	O07175 mycobacteri
8	989	26.8	355	16	Q7U2S9 mycobacteri
9	778.5	21.1	361	2	Q50320 mycobacteri
10	775.5	21.0	393	16	Q7IZJ3 mycobacteri
11	775.5	21.0	411	16	O53939 mycobacteri
12	758.5	20.6	410	2	Q99Q11 mycobacteri
13	754.5	20.5	354	16	Q9CCY9 mycobacteri
14	753.5	20.4	409	16	O53957 mycobacteri
15	753.5	20.4	409	16	Q7TZH7 mycobacteri
16	737	20.4	423	16	O53950 mycobacteri

17	737	20.0	423	16	Q7TZI4
18	736.5	20.0	399	16	Q7TZH8 mycobacteri
19	734	19.9	421	16	Q9Z5K0 mycobacteri
20	733.5	19.9	403	16	O53956 mycobacteri
21	700	19.0	413	16	O06386 mycobacteri
22	697	18.9	391	16	P96362 mycobacteri
23	697	18.9	391	16	P96362 mycobacteri
24	696	18.9	694	16	Q8VJW0 mycobacteri
25	689	18.7	468	16	O53958 mycobacteri
26	674	18.3	380	16	P95190 mycobacteri
27	674	18.3	380	16	Q7TX66 mycobacteri
28	671.5	18.2	394	16	Q7TXX5 mycobacteri
29	670.5	18.2	462	16	O33310 mycobacteri
30	667.5	18.1	385	16	Q7TZR7 mycobacteri
31	664.5	18.0	385	16	O33204 mycobacteri
32	663.5	18.0	385	16	Q8VJZ0 mycobacteri
33	641	17.4	363	16	O53940 mycobacteri
34	641	17.4	364	16	Q7TZJ5 mycobacteri
35	637	17.3	350	16	Q7TZJ2 mycobacteri
36	633.5	17.2	365	16	O86373 mycobacteri
37	633.5	17.2	405	16	Q8VJW5 mycobacteri
38	628.5	17.1	381	16	Q7TX67 mycobacteri
39	616.5	16.7	397	2	Q9AGF0 mycobacteri
40	615.5	16.7	402	16	O33312 mycobacteri
41	615.5	16.7	443	16	Q8VKL9 mycobacteri
42	613.5	16.6	382	16	Q7TXX3 mycobacteri
43	609.5	16.5	443	16	Q7U242 mycobacteri
44	605.5	16.4	423	16	Q7U114 mycobacteri
45	605.5	16.4	426	16	O05907 mycobacteri

ALIGNMENTS

RESULT 1

O05298	ID	O05298	PRELIMINARY;	PRT;	391 AA.
AC	005298;				
DT	01-JUL-1997 (TrEMBLrel. 04, Created)				
DT	01-JUL-1997 (TrEMBLrel. 04, Last sequence update)				
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)				
DE	Hypothetical protein (PPE family protein).				
GN	RV1196 OR MTC1364.08 OR MTC1234.				
OS	Mycobacterium tuberculosis.				
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;				
OC	Corynebacterineae; Mycobacteriaceae; Mycobacterium.				
OX	NCBI_TaxID=1773;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=H37RV;				
RX	MEDLINE=98295987; PubMed=9634230;				
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,				
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,				
RA	Davies R., Basham D., Brown D., Chillingworth T., Connor R.,				
RA	Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,				
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,				
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,				
RA	Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;				
RA	"Deciphering the biology of Mycobacterium tuberculosis from the				
RT	complete genome sequence."				
RN	Nature 393:537-544(1998).				
RC	SEQUENCE FROM N.A.				
RA	STRAIN=CDC 1551 / Oshkosh;				
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,				
RA	Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,				
RA	Kolonyak J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,				
RA	Deicher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,				
RA	Bishai W.;				
RT	"Whole genome comparison of Mycobacterium tuberculosis clinical and				
RT	laboratory strains."				
RL	Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.				


```

QY 322 LLEQAAVEEASDTAAANCLMNNVPOALQOQAQTOGTTTSSKLGKGLKWTSPHRSPIGN 381
Db 181 LLEQAAVEEASDTAAANCLMNNVPOALQOQAQTOGTTTSSKLGKGLKWTSPHRSPIGN 240
QY 382 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAQAVTAARQVTAONGVRAMSS-----LGSSL 437
Db 241 IVSMLNNHVMNTSGVSMNTLSSMLKGFAPAAA-QAVETAARQVTAONGVRAMSS-----LGSSL 299
QY 438 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAARPGQMGLG 497
Db 300 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAARPGQMGLG 359
QY 498 GLPVGQMGARAG--GGLSGVLRVPPPPYVPHSPAAAG 532
Db 360 GLPLGH-SVNNAGSGINNLRVPAEAYTPRTPAAG 393
QY 498 GLPVGQMGARAG--GGLSGVLRVPPPPYVPHSPAAAG 532
Db 360 GLPLGH-SVNNAGSGINNLRVPAEAYTPRTPAAG 393

RESULT 4
ID Q8VIZ3 PRELIMINARY; PRT; 393 AA.
AC Q8VIZ3;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DE 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE PPE family protein.
GN MT3582.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischnann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; A8007161; AAK47941.1; -.
DR TIGR; MT3582; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0008508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000030; Microbac_PPE.
DR InterPro; IPR000508; Peptidase_S26.
DR Pfam; PF00823; PPE; 1.
DR PROSITE; PS00501; SPASE I 1; 1.
SQ SEQUENCE 393 AA; 39688 MW; 86F0B67798855511 CRC64;

Query Match 44.7%; Score 1648; DB 16; Length 393;
Best Local Similarity 84.8%; Pred. No. 3.5e-71;
Matches 335; Conservative 20; Mismatches 34; Indels 6; Gaps 3;

QY 142 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201
Db 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 202 SSAGLMVAASPPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
Db 61 SSAGLMVAASPPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 262 LIATNLLGNTPTAIVNEAEYEGMWAQDAAMFGYAAATATATATLPPPEAPMTSAGG 321
Db 121 LIATNLLGNTPTAIVNEAEYEGMWAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180
QY 322 LLEQAAVEEASDTAAANCLMNNVPOALQOQAQTOGTTTSSKLGKGLKWTSPHRSPIGN 381
Db 181 LLEQAAVEEASDTAAANCLMNNVPOALQOQAQTOGTTTSSKLGKGLKWTSPHRSPIGN 240
QY 382 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAQAVTAARQVTAONGVRAMSS-----LGSSL 437
Db 241 IVSMLNNHVMNTSGVSMNTLSSMLKGFAPAAA-QAVETAARQVTAONGVRAMSS-----LGSSL 299
QY 438 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAARPGQMGLG 497

```

```

Db 241 IVSMLNNHVMNTSGVSMNTLSSMLKGFAPAAA-QAVETAARQVTAONGVRAMSS-----LGSSL 299
QY 438 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAARPGQMGLG 497
Db 300 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAARPGQMGLG 359
QY 498 GLPVGQMGARAG--GGLSGVLRVPPPPYVPHSPAAAG 532
Db 360 GLPLGH-SVNNAGSGINNLRVPAEAYTPRTPAAG 393
QY 498 GLPVGQMGARAG--GGLSGVLRVPPPPYVPHSPAAAG 532
Db 360 GLPLGH-SVNNAGSGINNLRVPAEAYTPRTPAAG 393

RESULT 5
ID O06341 PRELIMINARY; PRT; 393 AA.
AC O06341;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
DE 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Hypothetical protein RV3478.
GN RV3478 OR MTCY13E12.31.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RA MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean S., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Ruster S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL; Z95390; CAB08702.1; -.
DR PIR; C70568; C70568.
DR TubercuList; RV3478; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;

Query Match 42.9%; Score 1583; DB 16; Length 393;
Best Local Similarity 81.5%; Pred. No. 4.4e-68;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 142 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201
Db 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 202 SSAGLMVAASPPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 261
Db 61 SSAGLMVAASPPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 262 LIATNLLGNTPTAIVNEAEYEGMWAQDAAMFGYAAATATATATLPPPEAPMTSAGG 321
Db 121 LIATNLLGNTPTAIVNEAEYEGMWAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180
QY 322 LLEQAAVEEASDTAAANCLMNNVPOALQOQAQTOGTTTSSKLGKGLKWTSPHRSPIGN 381
Db 181 LLEQAAVEEASDTAAANCLMNNVPOALQOQAQTOGTTTSSKLGKGLKWTSPHRSPIGN 240
QY 382 MVSMAHHMNTSGVSMNTLSSMLKGFAPAAAQAVTAARQVTAONGVRAMSS-----LGSSL 437
Db 241 IVSMLNNHVMNTSGVSMNTLSSMLKGFAPAAA-QAVETAARQVTAONGVRAMSS-----LGSSL 299
QY 438 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAARPGQMGLG 497

```

Db 300 GSSGLGAGVAANLGRAASVGSLSVPPAWAANAQVTPAARALPLTSLTSAQAQPGHMLG 359
 Qy 498 GLPVQGMARAGGSLGVLRVPPRYPMHPSPAAG 532
 Db 360 GLPLGH-SVNAGSGINNLRVPRAYAIPTPTPAAG 393

RESULT 6
 Q7TWFS PRELIMINARY; PRT; 393 AA.
 AC Q7TWFS;
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
 DE PE family protein.
 GN PP560 OR M53505.
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12788972;
 RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Akin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
 RA "The complete genome sequence of Mycobacterium bovis";
 RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
 RL EMBL; BX248346; CAD95692.1; -.
 KW Complete proteome.
 SQ SEQUENCE 393 AA; 39413 MW; AF4C20C95DAE7DD4 CRC64;

Query Match 42.9%; Score 1583; DB 16; Length 393;
 Best Local Similarity 81.5%; Pred. No. 4.4e-68;
 Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

Qy 142 MVDFGALPPEINSARMYAGSGASLSVAAAQWMSVASDLFSAASAFQSVVWGLTVGSGWIG 201
 Db 1 MVDFGALPPEINSARMYAGSGASLSVAAAQWMSVASDLFSAASAFQSVVWGLTVGSGWIG 60
 Qy 202 SSAGLMAAASPYVWMSVTAQAGLTAQVRAAAAYETAYGLTVPVPIAENRAELMI 261
 Db 61 SSAGLMAAASPYVWMSVTAQAGLTAQVRAAAAYETAYGLTVPVPIAENRAELMT 120
 Qy 262 LIATNLLGONTPTAIVNEAEYGENWAQDAAMFYAAATATATATLPPPEAEPMTSAGG 321
 Db 121 LTATNLLGONTPTAIVNEAEYGENWAQDAAMFYAAATATATATLPPPEAEPMTSAGG 180
 Qy 322 LLEQAAVEEASDTAAANQIMNVPQALQLOAQPTQGTSSKLGKLTWVSPHSPISN 381
 Db 181 LLEQAAVEEASDTAAANQIMNVPQALQLOAQPTQGTSSKLGKLTWVSPHSPISN 240
 Qy 382 MVSANNHMTSGVNTNTLSSMLKGFAPAAARQAVQTAQAQNGVRAMS-----LCSSIL 437
 Db 241 VSSANNHMTSGVNTNTLSSMLKGFAPAAARQAVQTAQAQNGVRAMS-----LCSSIL 299
 Qy 438 GSSGLGGVAAANLGRAASVGSLSVPPAWAANAQVTPAARALPLTSLTSAARGPQMLG 497
 Db 300 GSSGLGAGVAANLGRAASVGSLSVPPAWAANAQVTPAARALPLTSLTSAQAQPGHMLG 359

Qy 498 GLPVQGMARAGGSLGVLRVPPRYPMHPSPAAG 532
 Db 360 GLPLGH-SVNAGSGINNLRVPRAYAIPTPTPAAG 393

RESULT 7
 Q07175 PRELIMINARY; PRT; 355 AA.
 AC Q07175;
 DT 01-JUL-1997 (TrEMBLrel. 04, Created)

DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Hypothetical protein (Serine protease, putative).
 GN PEPA OR RV0125 OR MTC1418B.07 OR MT0133.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Emdin M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL; Z96071; CAB09453.1; -.
 DR EMBL; AE006925; AAK44357.1; -.
 DR PIR; F70983; F70983.
 DR TIGR; MT0133; -.
 DR Tuberculin; RV0125; -.
 DR GO; GO:0004295; F:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys Ser trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase S1.
 DR InterPro; IPR001940; Peptidase_S1b_V8.
 DR Pfam; PF00595; PDZ; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Hypothetical protein; Serine protease; Protease;
 KW Complete proteome.
 SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match 26.8%; Score 989; DB 16; Length 355;
 Best Local Similarity 98.0%; Pred. No. 8.4e-40;
 Matches 196; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 530 AAGDIAPALSDRFPADFPALDPSAMVAQVQVNVNINIKLGYNNVAGTGTIVDPN 589
 Db 28 APAQAAPALSDRFPADFPALDPSAMVAQVQVNVNINIKLGYNNVAGTGTIVDPN 87
 Qy 590 GVLTNHVIAGATDINAFSGVSGTGYVDVYVDRTQDVAVLQIRGAGGLPSAIGGV 649
 Db 88 GVLTNHVIAGATDINAFSGVSGTGYVDVYVDRTQDVAVLQIRGAGGLPSAIGGV 147
 Qy 650 AVGEPVWAMGSGGGTTPRVPGRVVALGQVQASDLSLTGAETLNGLIQDAAIPGD 709

Db 148 AVGEPVVMGNSGGQGTTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGD 207
QY 710 SGGPVVNGLGQVVGMMNTAAS 729
Db 208 SGGPVVNGLGQVVGMMNTAAS 227

RESULT 8
QYU2S9 PRELIMINARY; PRT; 355 AA.
ID QYU2S9
AC QYU2S9
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable serine protease PEPA (EC 3.4.21.-).
GN PEPA OR MB0130.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=AF2122/97;
RC MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL: BX248334; CAD92991.1; -.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match 26.8%; Score 989; DB 16; Length 355;
Best Local Similarity 98.0%; Pred. No. 8.4e-40;
Matches 196; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 530 AAGIAPALSDPFAFPALPDPSSAWAQVGPQVNNIKLGYNNVAGAGTIVDPN 589
Db 28 APAQAAPALSDPFAFPALPDPSSAWAQVGPQVNNIKLGYNNVAGAGTIVDPN 87

QY 590 GWLITNNHVIAGATDINAFSGSGQTYGVDVYDRTQDVAVLQRLGAGGLPSAAIGGV 649
Db 88 GWLITNNHVIAGATDINAFSGSGQTYGVDVYDRTQDVAVLQRLGAGGLPSAAIGGV 147

QY 650 AVGEPVVMGNSGGQGTTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGD 709
Db 148 AVGEPVVMGNSGGQGTTPRAVGRVVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGD 207

QY 710 SGGPVVNGLGQVVGMMNTAAS 729
Db 208 SGGPVVNGLGQVVGMMNTAAS 227

RESULT 9
QYU320 PRELIMINARY; PRT; 361 AA.
ID QYU320
AC QYU320
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 34kDa protein precursor.
OS Mycobacterium paratuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1770;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=JD88/107;
RC MEDLINE=95005449; PubMed=7921248;
RA Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;
RT "Identification and characterisation of a putative serine protease

expressed in vivo by Mycobacterium avium subsp paratuberculosis.";
RL Microbiology 140:1977-1982(1994).
CC SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
DR EMBL: Z23092; CRA80638.1; -.
DR PIR: S47170; S47170.
DR GO: GO:0008233; F:peptidase activity; IEA.
DR GO: GO:0004295; F:trypsin activity; IEA.
DR GO: GO:0007242; P:intracellular signaling cascade; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ; 1.
DR PRINTS: PR00834; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS50106; PDZ; 1.
DR Hydrolase; Protease; Serine protease; Signal.
FT SIGNAL 38 POTENTIAL.
SQ SEQUENCE 361 AA; 35709 MW; 30FEP78FD6F3C411 CRC64;

Query Match 21.1%; Score 778.5; DB 2; Length 361;
Best Local Similarity 66.0%; Pred. No. 9.5e-30;
Matches 155; Conservative 28; Mismatches 37; Indels 15; Gaps 2;

QY 495 MGGGLPVGMGARAGGGLSGVLRVPRPYVMPHSPAGDIAPALSDPFAFPALPDP 554
Db 15 LVGVLTVVGLGLGLSGVG-----LAPASA--APSGLALDRFADRLAPIDP 59

QY 555 SAMVAQVGPQVNNIKLGYNNVAGAGTIVDPNGVLTNNHVIAGATDINAFSGSGQ 614
Db 60 SAMVQVGPQVNNIKLGYNNVAGAGTIVDPNGVLTNNHVIAGATDINAFSGSGQ 119

QY 615 TYGVVGVYDRTQDVAVLQRLGAGGLPSAAIGGVAVGVPVVMGNSGGQGTTPRAVGR 674
Db 120 TVAVDVGVYDRTQDVAVLQRLGAGGLPSAAIGGVAVGVPVVMGNSGGQGTTPRAVGR 179

QY 675 VVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 729
Db 180 VVALGQTVQASDSLTGAEETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 234

RESULT 10
QYU7J3 PRELIMINARY; PRT; 393 AA.
ID QYU7J3
AC QYU7J3
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PPE family protein.
GN PPE26 OR M51817.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=AF2122/97;
RC MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis."
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
DR EMBL: BX248340; CAD94520.1; -.
KW Complete proteome.
SQ SEQUENCE 393 AA; 38572 MW; 41944B6E547ABAB0 CRC64;

Query Match 21.0%; Score 775.5; DB 16; Length 393;
Best Local Similarity 43.4%; Pred. No. 1.5e-29;
Matches 178; Conservative 65; Mismatches 130; Indels 37; Gaps 10;

DR TubercList: Rv1789; -
 DR InterPro: IPR000030; Microbac_ppe.
 DR Pfam: PF00823; PPE; 1.
 DR Complete proteome.
 SQ SEQUENCE 411 AA; 40557 MW; 03C90B5E0590B7DA CRC64;

Query Match 21.0%; Score 775.5; DB 16; Length 411;
 Best Local Similarity 43.4%; Pred. No. 1.6e-29;
 Matches 178; Conservative 65; Mismatches 130; Indels 37; Gaps 10;

QY 143 VDFGALPPEINSARMYAGPGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGSWIGS 202
 DB 1 MDFGALPPEVNSVRMTYAGPGSAPMWAAASAWNGLAELSSAATGYETVITQLSSSGWGLGP 60

QY 203 SAGLMVAASPPYVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPIAENRAELMIL 262
 DB 61 ASAMAEEAVAPYVAMSVAAAQAQAEQATQARAAAAFEAFATVPPPLIAANRASLMQL 120

QY 263 IATNLLGONTPALAVNEAEYGENWQAQDAAMFGYAAATATATATLPPFEAPEMTSAGGL 322
 DB 121 ISTNVGQNTSALAAAEQYGENWQAQDSAAAYAGSSASASA-VTPFSTPQIANPTAQ 179

QY 323 LEQAAAEVAAEDTAAA--NOLMNVVQALQQLAQTQGTTPSSKLGGLWK----- 370
 DB 180 GTQAAAVATAAGTAQSTLTMITGLENALQSLTSPQLQSS--NGPLSLWMLQILFGTNPFT 238

QY 371 -----TVSPHRSPISNMVMNMMNTSGVSMNTLSSMLKGFAPAAARQAVQAAA 423
 DB 239 SISALLTDLPQYASFPYNTGELPYFSGIGNNFIQAATL-GLIGSAAFAAV-----AAA 292

QY 424 QNGVRAWSLGSGLSGSGGGVAAANLGRAASVGSLSVQAWAA--ANQAVTPAARALPLT 482
 DB 293 GDAAKGLPGLGMLG----GGPVAAGLGNAAVSGKLSVPPVWSGGLPGSVTPGAAPLPVS 348

QY 483 SLTSAARPGQVGLGGLPVQMGARAGGGLSGVLVPPRPYVMPHSPAAG 532
 DB 349 TVSAAPEAAPGSLGGLPL----AGAGGAGAGP-RYGFRPTVMARPPFFAG 393

RESULT 11
 OS3939 PRELIMINARY; PRT; 411 AA.
 AC OS3939;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE PPE-family protein.
 GN Rv1789 OR MT1838 OR MT049.11.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]_TaxID=1773;
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37Rv;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jørgensen K., Krogh A., McClean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skellon S., Squares R., Squares R.,
 RA Sultston J.E., Taylor K., Whitehead S., Barrett B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence";
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Kouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; AL022021; CAAL17711.1; ALT_INIT.
 DR EMBL; AE007043; AAK46108.1; -;
 DR PIR; G70929; G70929.
 DR TIGR; MT1838; -;

DR TubercList: Rv1789; -
 DR InterPro: IPR000030; Microbac_ppe.
 DR Pfam: PF00823; PPE; 1.
 DR Complete proteome.
 SQ SEQUENCE 411 AA; 40557 MW; 03C90B5E0590B7DA CRC64;

Query Match 21.0%; Score 775.5; DB 16; Length 411;
 Best Local Similarity 43.4%; Pred. No. 1.6e-29;
 Matches 178; Conservative 65; Mismatches 130; Indels 37; Gaps 10;

QY 143 VDFGALPPEINSARMYAGPGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGSWIGS 202
 DB 19 MDFGALPPEVNSVRMTYAGPGSAPMWAAASAWNGLAELSSAATGYETVITQLSSSGWGLGP 78

QY 203 SAGLMVAASPPYVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPIAENRAELMIL 262
 DB 79 ASAMAEEAVAPYVAMSVAAAQAQAEQATQARAAAAFEAFATVPPPLIAANRASLMQL 138

QY 263 IATNLLGONTPALAVNEAEYGENWQAQDAAMFGYAAATATATATLPPFEAPEMTSAGGL 322
 DB 139 ISTNVGQNTSALAAAEQYGENWQAQDSAAAYAGSSASASA-VTPFSTPQIANPTAQ 197

QY 323 LEQAAAEVAAEDTAAA--NOLMNVVQALQQLAQTQGTTPSSKLGGLWK----- 370
 DB 198 GTQAAAVATAAGTAQSTLTMITGLENALQSLTSPQLQSS--NGPLSLWMLQILFGTNPFT 256

QY 371 -----TVSPHRSPISNMVMNMMNTSGVSMNTLSSMLKGFAPAAARQAVQAAA 423
 DB 257 SISALLTDLPQYASFPYNTGELPYFSGIGNNFIQAATL-GLIGSAAFAAV-----AAA 310

QY 424 QNGVRAWSLGSGLSGSGGGVAAANLGRAASVGSLSVQAWAA--ANQAVTPAARALPLT 482
 DB 311 GDAAKGLPGLGMLG----GGPVAAGLGNAAVSGKLSVPPVWSGGLPGSVTPGAAPLPVS 366

QY 483 SLTSAARPGQVGLGGLPVQMGARAGGGLSGVLVPPRPYVMPHSPAAG 532
 DB 367 TVSAAPEAAPGSLGGLPL----AGAGGAGAGP-RYGFRPTVMARPPFFAG 411

RESULT 12
 Q99Q11 PRELIMINARY; PRT; 410 AA.
 AC Q99Q11;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Rv1808-like protein.
 GN MYC1808 OR Ovi1808.
 OS Mycobacterium microti.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1806;
 RN [1]_TaxID=1806;
 RP SEQUENCE FROM N.A.
 RC STRAIN=MYC 94-2272, and OV254;
 RA Sivodon V., Heym B., Mazancourt P., Gaillard J.-L.;
 RT "PPE Rv1808 orthologue of Mycobacterium microti";
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBSJ databases.
 DR EMBL; AF335179; AAK20893.1; -;
 DR EMBL; AF335179; AAK20893.1; -;
 DR InterPro; IPR000030; Microbac_ppe.
 DR Pfam; PF00823; PPE; 1.
 SQ SEQUENCE 410 AA; 40044 MW; 3B9DA3174655A5EA CRC64;

Query Match 20.6%; Score 758.5; DB 2; Length 410;
 Best Local Similarity 41.5%; Pred. No. 1e-28;
 Matches 183; Conservative 51; Mismatches 142; Indels 65; Gaps 10;

QY 142 MVDFGALPPEINSARMYAGPGSASLVAAQWMDVSADLFSASAFQSVVWGLTVGSWIG 201
 DB 1 MLDGALPPEVNSVRMTYAGPGSAPMWAAASAWNGLAELSSAATGYETVITQLSSSGWGLGP 60

QY 202 SSAGLMVAASPPYVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPIAENRAELMI 261

Db 61 PSSITMAAAVAVYAWISVTAGQAEQAQAQAKIAAGVYETAFAPVPPVIEANRALIMS 120
QY 262 LIATNLGONTFAIAVNEAEVGEWAQDAAMFGYAAATATATATATALLPFEAPETSAGG 321
Db 121 LVATNIFGONTFAIAATEAHYAEWAQDAAMFGYAGSATA-SQAFPSPPQTTFESA 179
QY 322 LLEQAAAEVEASDTAAA-----NQLMNVNPOALQQLAQFTQGTTPSSKLGGLWKTVP 374
Db 180 TAAQSAVVAQAAGAAASSDITAQLSQLSLPSTLQSLA--TTATATSASAG--WDTVL- 234
QY 375 HRSPTISNMVSMANNHSMNTNSGVMSNTLSSMLKGFAPAA-----ARQAVQTAA 423
Db 235 -----QSITILANLTGPYSIIIGLGAIPGGWMLTFGQILGLAQNAPGVAA 279
QY 424 QNGVRAVMSLGSGLGS-----SGLGGVAAVNLGRAASVGSLSVFPQAAANQAQVTP 474
Db 280 LLGPRAAGALSPLAPLRGGYADITPLGGGATGGIAIYVGSLSVFPQAAEAAPVRA 339
QY 475 ARAALPLTSLTA-AERPGQMLGLPVQMGARAGGLJ-----SGVLRVPRFPVPMHS 528
Db 340 VASVLPFGTGAAPALAAEAPGALFGEMALSSLAGRALAGTAVRSAGAAARV----- 389
QY 529 PAAGDIAPPALSQDRFADPPA 549
Db 390 -AGGSVTEADVASTTIIIVIPA 409

RESULT 13

ID Q9CCV9 PRELIMINARY; PRT; 354 AA.
AC Q9CCV9;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable secreted serine protease.
GN ML2659.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -1- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL583926; CAC32191.1; -;
DR PIR; A87242; 287242.
DR Leproma; M2659; -;
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys Ser trypsin.
DR InterPro; IPR001478; PDZ_
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR008256; Peptidase_S1b_v8.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SM00228; PDZ; 1.

DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 354 AA; 35265 MW; 612F23261BC9EA4A CRC64;
Query Match 20.5%; Score 754.5; DB 16; Length 354;
Best Local Similarity 73.3%; Pred. No. 1.3e-28;
Matches 151; Conservative 25; Mismatches 25; Indels 5; Gaps 2;
QY 524 VMPHSPAAGDIAPPALSQDRFADFPALPLDPSAMVAQVGPQVYVNTKLGYNNAVGAGTG 583
Db 27 VVFGS--ATSPGSLTALDRFSNRPLPLNPAAMVA---PQVNVISRLCYNSAVGAGTG 81
QY 584 IVIDPNGVLTNNHVIAGATDINAFVSGQTYGVVDVWGYDRTQDYAVLQLRGAGGLPSA 643
Db 82 IVIDSSGWLVTNNHVISGATDISAFDVGNGKTYGVVDVWGYDRTQDYAVLQLRGAGGLPTA 141
QY 644 AIGGVAVGPPVVMGNSGGGCTPRAVPGRVVALGQTVQASDLSLGAETLNGLIQFDA 703
Db 142 VIGDVAIGPIVALGNTGGGSLPSVLPGRVVALNQTVQASEPLTGAQETLSGLIQVDA 201
QY 704 AIOPGDSGGPVVNLGGOVGMNTAAS 729
Db 202 PIKPGDSGGPVVNSRGQVGMNTAAT 227

RESULT 14

ID O53957 PRELIMINARY; PRT; 409 AA.
AC O53957;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PPE-family protein.
GN RV1808 OR MT1856.1 OR MTU049.30.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., Maclean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Emdolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022021; CAAL17729.1; -;
DR EMBL; AE007044; AAK46129.1; ALT_INIT.
DR PIR; A70932; A70932.
DR TIGR; MT1856.1; -;
DR TubercuList; RV1808; -;
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Complete proteome.
SQ SEQUENCE 409 AA; 39917 MW; 1E15202BACF36379 CRC64;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:39:00 ; Search time 74.9614 Seconds
(without alignments) 2747.774 Million cell updates/sec

Title: US-09-597-796C-12
Perfect score: 3686
Sequence: 1 MHHHHHTAASDNFOLSQGG.....SGGPVNGIGQVVGMMNTAAS 729

Scoring table: BLOSUM62

Searched: 1586107 seqs. 282547505 residues

total number of hits satisfying	chosen parameters:	1586107
---------------------------------	--------------------	---------

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%

Maximum March 1999
Listing first 45 summaries

Database : A Geneseq 29Jan04:

- 1: geneseqp1980s: *
- 2: geneseqp1990s: *
- 3: geneseqp2000s: *
- 4: geneseqp2001s: *
- 5: geneseqp2002s: *
- 6: geneseqp2003as: *
- 7: geneseqp2003bs: *
- 8: geneseqp2004s: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	3686	100.0	729	5	AAE29708	Mycobacte
2	3686	100.0	729	5	AAE17572	Mycobacte
3	3680	99.8	729	4	AAO21442	Ra12-H9-3
4	3680	99.8	813	7	ADA26367	Mycobacte
5	3680	99.8	825	7	ADA26366	Mycobacte
6	3680	99.8	875	7	ADA26365	Mycobacte
7	3680	99.8	930	5	AAE29731	Mycobacte
8	3680	99.8	930	7	ADA26364	Mycobacte
9	3680	99.8	1016	7	ADA26370	M. bovis
10	3680	99.8	1022	7	ADA26369	Mycobacte
11	3680	99.8	1154	7	ADA26368	Mycobacte
12	3677	99.8	729	5	AAE29709	Mycobacte
13	3677	99.8	729	5	AAE17573	Mycobacte
14	3677	99.8	729	7	ADA26374	Mycobacte
15	3670	99.6	729	7	ADA26373	Mycobacte
16	3657	99.2	729	2	AAI32059	Mycobacte
17	3628.5	98.4	726	5	AAU74588	Antigenic
18	2941	79.8	596	2	AAI32070	Mycobacte
19	2941	79.8	596	5	AAE29710	Mycobacte
20	2941	79.8	595	5	AAE17574	Mycobacte
21	2941	79.8	599	5	AAU74599	Antigenic
22	2637	71.5	1010	7	ADA26356	Mycobacte
23	2631	71.4	723	7	ADA26354	Mycobacte
24	2002	54.3	788	4	AAU01903	M. tuberc
25	2001.5	54.3	744	4	AAU01902	M. tuberc

26	2001.5	54.3	815	4	AAU01904	AAU01904	M. tuberc
27	1967.5	53.4	600	2	AA32068	AA32068	Mycobacte
28	1967.5	53.4	600	5	AAU74597	AAU74597	Antigenic
29	1944	52.7	391	2	AAW32381	AAW32381	Mycobacte
30	1944	52.7	391	2	AAW32449	AAW32449	Mycobacte
31	1944	52.7	391	2	AAW64335	AAW64335	Mycobacte
32	1944	52.7	391	2	AAW81702	AAW81702	M. tuberc
33	1944	52.7	391	2	AAW04778	AAW04778	Mycobacte
34	1944	52.7	391	2	AA338989	AA338989	M. tuberc
35	1944	52.7	391	2	AA339132	AA339132	M. tuberc
36	1944	52.7	391	4	AAU01988	AAU01988	M. tuberc
37	1944	52.7	391	5	AAE29707	AAE29707	Mycobacte
38	1944	52.7	391	5	AAE17571	AAE17571	Mycobacte
39	1939	52.6	394	2	AAW04779	AAW04779	Mycobacte
40	1652.5	44.8	396	2	AAW64337	AAW64337	Mycobacte
41	1652.5	44.8	396	2	AAW81704	AAW81704	M. tuberc
42	1652.5	44.8	396	2	AA338991	AA338991	M. tuberc
43	1652.5	44.8	396	2	AA339134	AA339134	M. tuberc
44	1593	42.9	393	6	ABU36968	ABU36968	Protein e
45	1593	42.9	393	6	ABU34531	ABU34531	Protein e

ALIGNMENTS

RESULT 1	
AAE29708	
ID	AAE29708 standard; protein; 729 AA.
XX	XX
AC	AAE29708;
XX	XX
XX	29-AUG-2003 (revised)
DT	27-JAN-2003 (first entry)
DT	XX
XX	Mycobacterium sp. MTB72F fusion protein.
DE	XX
XX	Vaccine; immunity; diagnostic agent; gene
KW	Ra12; MTB72F; fusion protein.
KW	XX
XX	Mycobacterium sp.
OS	Mycobacterium tuberculosis.
OS	Chimeric.
XX	XX
XX	WO200272792-A2.
FN	XX
XX	19-SEP-2002.
XX	XX
XX	13-MAR-2002: 2002WO-US008223.
XX	XX

XX Skeiky Y, Brannon M, Guderian J;
PI
XX
DR WPI; 2002-759844/82.
DE N-PSDB: AAD47083.

XX New recombinant nucleic acid molecule comprising a Leishmania TSA, LeIF,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 PT tuberculosis.

Page 90-92; 155pp. English.

XX CC The invention relates to a recombinant nucleic acid molecule encoding a
CC CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
CC CC polynucleotide sequence encoding an antigen or an antigenic fragment from
CC CC *Mycobacterium* sp. and a *Leishmania* polynucleotide sequence encoding a
CC CC polypeptide or its fragment. The *Leishmania* polynucleotide is selected
CC CC from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
CC CC are used in methods for eliciting immune response in mammals. They are
CC CC useful as vaccines to elicit protective immunity against pathogenic

CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
CC polypeptides are used for enhancing the expression of polynucleotides, as
CC in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is
CC MTB72P fusion protein. This fusion protein comprises Ral2 and Ra35
CC protein from M. tuberculosis and TBH9 protein from Mycobacterium sp.
CC (Updated on 29-AUG-2003 to standardise OS field)
XX
SQ Sequence 729 AA;

Query Match 100.0%; Score 3686; DB 5; Length 729;
Best Local Similarity 100.0%; Pred. No. 1.3e-233; Indels 0; Gaps 0;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHTAASDNFOLSGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVD 60
Db 1 MHHHHHTAASDNFOLSGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVD 60
QY 61 NNGNGARVORVVGSAAPASLSIGTGDVITAVDGPINSATAMADALNGHHGPDVISTWQ 120
Db 61 NNGNGARVORVVGSAAPASLSIGTGDVITAVDGPINSATAMADALNGHHGPDVISTWQ 120
QY 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
Db 121 TKSGGTRTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
QY 181 FSAASAFQSVWGLTVGWSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
Db 181 FSAASAFQSVWGLTVGWSAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
QY 241 TAYGLTVPPVTAENRAELMILIAITLLGONTPALAVNEAEYGEWAOADRAAMGYAAT 300
Db 241 TAYGLTVPPVTAENRAELMILIAITLLGONTPALAVNEAEYGEWAOADRAAMGYAAT 300
QY 301 ATATATLPPFEPAPMTSAGGLLEQAAAEEASDTAAANQLMNVYQALQLAQPTQTT 360
Db 301 ATATATLPPFEPAPMTSAGGLLEQAAAEEASDTAAANQLMNVYQALQLAQPTQTT 360
QY 361 PSSKLGGLWKTSPHRSPTISNNVSMANNHMTNSGVSMNTLSMLKGFAPAARQAVQ 420
Db 361 PSSKLGGLWKTSPHRSPTISNNVSMANNHMTNSGVSMNTLSMLKGFAPAARQAVQ 420
QY 421 TAAQNGVRAMSSGLSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALP 480
Db 421 TAAQNGVRAMSSGLSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALP 480
QY 481 LTSLSAABERGPGQMLGGLPVQGMGARAGGGLSGVLVPPRPYVPHSPHSPAAAGDIAP 540
Db 481 LTSLSAABERGPGQMLGGLPVQGMGARAGGGLSGVLVPPRPYVPHSPHSPAAAGDIAP 540
QY 541 QDRFADFPALPLDPSAMVAQVGPVNTKLGYNNAVAGTGIVIDPNGVLLTNHNVIA 600
Db 541 QDRFADFPALPLDPSAMVAQVGPVNTKLGYNNAVAGTGIVIDPNGVLLTNHNVIA 600
QY 601 GATDINAFSVSGGQTYGVVYVDRDQDAVQLRGAGGLPSAAIGGGVAVGEPVWAMGN 660
Db 601 GATDINAFSVSGGQTYGVVYVDRDQDAVQLRGAGGLPSAAIGGGVAVGEPVWAMGN 660
QY 661 SGGGGTTPRVPGRVVALGQTVQASDSITGAEETLNGLIQFDAIQDSDGGPVNGLGQ 720
Db 661 SGGGGTTPRVPGRVVALGQTVQASDSITGAEETLNGLIQFDAIQDSDGGPVNGLGQ 720
QY 721 VVGMTAAS 729
Db 721 VVGMTAAS 729

RESULT 2
AAE17572 standard; protein; 729 AA.
XX
AC
XX

DT 22-APR-2002 (first entry)
XX Mycobacterium species MTB72F fusion protein.
DE
XX Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; MTB72F; Ral2-TbH9-Ra35 protein.
XX
OS Mycobacterium sp.
XX
FH Key Location/Qualifiers
FT Region 8..139
FT Region /label= Ral2_protein
FT Region 142..532
FT Region /label= TbH9PL_protein
FT Region 535..729
FT Region /label= Ra35_protein
XX
FN W0200198460-A2.
XX
XX 27-DEC-2001.
XX
XX 20-JUN-2001; 2001WO-US019959.
PP
XX 20-JUN-2000; 2000US-00597796.
PR
XX 01-FEB-2001; 2001US-0265737P.
PR
XX (CORI-) CORIXA CORP.
PA
XX
XX Skeiky Y, Reed S, Alderson M;
PI
XX WPI; 2002-147798/19.
DR
XX N-PSDB; AAD28342.
DR
XX
PT Composition comprising MTB39 antigen and MTB32A antigen from
Myobacterium species, useful for eliciting immune response in a subject.
XX
XX Claim 6; Fig 5; 136pp; English.
XX
CC The present invention relates to fusion proteins containing at least two
CC Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected with
CC Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is a DNA encoding
CC Mycobacterium species MTB72F (Ral2-TbH9-Ra35) fusion protein
SQ Sequence 729 AA;

Query Match 100.0%; Score 3686; DB 5; Length 729;
Best Local Similarity 100.0%; Pred. No. 1.3e-233;
Matches 729; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MHHHHHTAASDNFOLSGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVD 60
Db 1 MHHHHHTAASDNFOLSGGQGFAPIGQAMAIAGQIRSGGSPVTHIGPTAFGLGLGVVD 60
QY 61 NNGNGARVORVVGSAAPASLSIGTGDVITAVDGPINSATAMADALNGHHGPDVISTWQ 120
Db 61 NNGNGARVORVVGSAAPASLSIGTGDVITAVDGPINSATAMADALNGHHGPDVISTWQ 120

QY 121 TKSGGTRTGNVTLAGPPAEFVWDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
 Db 121 TKSGGTRTGNVTLAGPPAEFVWDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
 QY 181 FSAASAFQSVVWGLTVGWSGAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
 Db 181 FSAASAFQSVVWGLTVGWSGAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
 QY 241 TAYGLTVPPVIAENRAELMILITATNLGONTPTAVNEAEYGEWMAQDAAMFGYAAAT 300
 Db 241 TAYGLTVPPVIAENRAELMILITATNLGONTPTAVNEAEYGEWMAQDAAMFGYAAAT 300
 QY 301 ATATATLLPFEAPMTSAGGLLEQAAAVVEASDTAAANQLMNNVPQALQLAQPTGTT 360
 Db 301 ATATATLLPFEAPMTSAGGLLEQAAAVVEASDTAAANQLMNNVPQALQLAQPTGTT 360
 QY 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMSMTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
 Db 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMSMTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
 QY 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
 Db 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
 QY 481 LTSLSAAERGPQMLGGLPVGQMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
 Db 481 LTSLSAAERGPQMLGGLPVGQMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
 QY 541 QDRFADFPALPLDPSAMVAQVGPQVNVNTKLGYNNAVAGAGTGVIPDNGVLTNNHVA 600
 Db 541 QDRFADFPALPLDPSAMVAQVGPQVNVNTKLGYNNAVAGAGTGVIPDNGVLTNNHVA 600
 QY 601 GATDINAFSVGSGQTVGVDDVVGVDYRDTQDVAVLQLRGAGGLPSAAITGGGVAVGEPVAVGN 660
 Db 601 GATDINAFSVGSGQTVGVDDVVGVDYRDTQDVAVLQLRGAGGLPSAAITGGGVAVGEPVAVGN 660
 QY 661 SGQGGTTPRAVGRVVALGQTVQASDSLTAEBETLNGLIQFDAATOPGDSGPPVNGLGQ 720
 Db 661 SGQGGTTPRAVGRVVALGQTVQASDSLTAEBETLNGLIQFDAATOPGDSGPPVNGLGQ 720
 QY 721 VVGMTAAS 729
 Db 721 VVGMTAAS 729
 RESULT 3
 ID AAO22142
 XX AAO22142 standard; protein; 729 AA.
 AC AAO22142;
 DT 03-OCT-2002 (first entry)
 XX
 DE Ra12-H9-32A fusion protein.
 XX
 KW Ra12; serine protease antigen; MTB32A; Mycobacterium tuberculosis;
 KW vaccine; eukaryotic; prokaryotic origin; serum antibody; fusion protein;
 KW immunogen; cytokines.
 XX
 OS Unidentified.
 OS Chimeric.
 PN WO200125401-A2.
 XX
 PD 12-APR-2001.
 XX
 PF 06-OCT-2000; 2000WO-US027652.
 XX
 PR 07-OCT-1999; 99US-015858SP.
 XX
 PA (CORI-) CORIXA CORP.
 XX

PI Skeiky Y, Guderian J;
 DR WPI: 2001-266299/27.
 DR N-PSDB; AAL40773.
 XX
 PT Recombinant nucleic acid molecule for producing high yield expression of
 PT desired fusion polypeptides, encodes fusion polypeptide comprising
 PT Mycobacterium tuberculosis coding sequence and heterologous polypeptide.
 XX
 PS Disclosure; Fig 6; 39pp; English.
 XX
 CC The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide, comprising a polynucleotide sequence of Ra12, a 14
 CC kDa C-terminal fragment of serine protease antigen MTB32A of
 CC Mycobacterium tuberculosis, and a heterologous polynucleotide sequence.
 CC The recombinant fusion nucleic acids and polypeptides are useful for
 CC providing stable and high yield expression of fusion polypeptides of both
 CC eukaryotic and prokaryotic origin and to encode a protein product for use
 CC as an antigen for detecting serum antibodies. The presence of serum
 CC antibodies to M. tuberculosis antigens in an individual indicates that
 CC the individual is infected with it. The fusion polypeptides are useful as
 CC sources of proteins for monitoring binding of serum antibodies to fusion
 CC proteins and as an immunogen to induce and/or enhance immune responses.
 CC The coding sequences can be ligated with a coding sequence of another
 CC molecule such as M. tuberculosis antigens, a cytokine or an adjuvant and
 CC can be used in vivo as a DNA vaccine. This sequence represents the Ra12-
 CC H9-32A fusion protein
 XX

SQ Sequence 729 AA;

Query Match 99.8%; Score 3680; DB 4; Length 729;
 Best Local Similarity 99.9%; Pred. No. 3.1e-233;
 Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHEHHHTAASDNFQLSQGGQGFAPITGQAWAAGAIRSGGSPYHIGFTAFGLGVD 60
 Db 1 MHHEHHHTAASDNFQLSQGGQGFAPITGQAWAAGAIRSGGSPYHIGFTAFGLGVD 60
 QY 61 NNGNGARVQVWGSAPAAASLGISTGVDITAVDCAPIINSATAMADALNGHHPGDIVSTWQ 120
 Db 61 NNGNGARVQVWGSAPAAASLGISTGVDITAVDCAPIINSATAMADALNGHHPGDIVSTWQ 120
 QY 121 TKSGGTRTGNVTLAGPPAEFVWDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
 Db 121 TKSGGTRTGNVTLAGPPAEFVWDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
 QY 181 FSAASAFQSVVWGLTVGWSGAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
 Db 181 FSAASAFQSVVWGLTVGWSGAGLMVAASPYVAMSVTAGQAEHTAAQVRVAAAAYE 240
 QY 241 TAYGLTVPPVIAENRAELMILITATNLGONTPTAVNEAEYGEWMAQDAAMFGYAAAT 300
 Db 241 TAYGLTVPPVIAENRAELMILITATNLGONTPTAVNEAEYGEWMAQDAAMFGYAAAT 300
 QY 301 ATATATLLPFEAPMTSAGGLLEQAAAVVEASDTAAANQLMNNVPQALQLAQPTGTT 360
 Db 301 ATATATLLPFEAPMTSAGGLLEQAAAVVEASDTAAANQLMNNVPQALQLAQPTGTT 360
 QY 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMSMTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
 Db 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMSMTNSGVSMNTLSSMLKGFAPAAARQAVQ 420
 QY 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
 Db 421 TAAQNGVRAMSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
 QY 481 LTSLSAAERGPQMLGGLPVGQMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
 Db 481 LTSLSAAERGPQMLGGLPVGQMGARAGGGLSGVLRVPPRYVMPHSPAAGDIAPPALS 540
 QY 541 QDRFADFPALPLDPSAMVAQVGPQVNVNTKLGYNNAVAGAGTGVIPDNGVLTNNHVA 600
 Db 541 QDRFADFPALPLDPSAMVAQVGPQVNVNTKLGYNNAVAGAGTGVIPDNGVLTNNHVA 600

QY 601 GATDINAFSVGGQTYGVYVVDYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGPPVAMGN 660
 DB 601 GATDINAFSVGGQTYGVYVVDYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGPPVAMGN 660
 QY 661 SGGGGTTPRAVPRGVAVGQTVQASDSLTGAETLNGLIQFDAAIOPGSGGPPVNGLGQ 720
 DB 661 SGGGGTTPRAVPRGVAVGQTVQASDSLTGAETLNGLIQFDAAIOPGSGGPPVNGLGQ 720
 QY 721 VVGNTAAS 729
 DB 721 VVGNTAAS 729

RESULT 4
 ADA26367
 ID ADA26367 standard; protein; 813 AA.
 AC ADA26367;
 DT 20-NOV-2003 (first entry)
 DE Mycobacterium MTB72F-DpV (fusion MTB81F) protein.
 KW fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 KW tuberculosis; tuberculostatic; gene therapy; vaccine.
 OS Chimeric.
 OS Mycobacterium sp.
 PN W02003070187-A2.
 XX W02003070187-A2.
 PD 28-AUG-2003.
 PF 18-FEB-2003; 2003WO-US004903.
 PR 15-FEB-2002; 2002US-0357351P.
 XX (CORI-) CORIXA CORP.
 PA Skeiky Y, Guderian J, Reed S;
 PI WPI; 2003-697554/66.
 DR N-PSDB; ADA26360.
 XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
 PT and MTB39 antigens, with or without the MTB85A antigen, from a
 PT Mycobacterium species, useful for diagnosing, preventing and/or treating
 PT tuberculosis infection.
 XX Claim 85; Fig 15; 112pp; English.
 PS The invention relates to a novel nucleic acid encoding a fusion
 CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
 CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
 CC A polypeptide of the invention has tuberculostatic activity. A
 CC polynucleotide of the invention may have a use in gene therapy, and as a
 CC vaccine. The methods and compositions of the present invention are useful
 CC for diagnosing, preventing and/or treating tuberculosis infection. The
 CC present sequence is used in the exemplification of the invention.
 XX Sequence 813 AA;
 SQ Query Match 99.8%; Score 3680; DB 7; Length 813;
 Best Local Similarity 99.9%; Pred. No. 3.6e-233;
 Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHTHHTAASDNFQISQGGGFAIPGQAMAIAGQIRSGGSPVHIGFTAFGLGVVD 60
 DB 1 MHHTHHTAASDNFQISQGGGFAIPGQAMAIAGQIRSGGSPVHIGFTAFGLGVVD 60
 QY 61 NNGNGARVQRVVGSAAPASLGISTGDVITAVDGPAINSATAMADALNGHHPGDIVSVTWQ 120

DB 61 NNGNGARVQRVVGSAAPASLGISTGDVITAVDGPAINSATAMADALNGHHPGDIVSVTWQ 120
 QY 121 TKSGETRTGNVTLAEGPPAEFMVDFGALPPEINSARMYAGFGSASLVAAQMWDSVASDL 180
 DB 121 TKSGETRTGNVTLAEGPPAEFMVDFGALPPEINSARMYAGFGSASLVAAQMWDSVASDL 180
 QY 181 FSAASAFQSVVWGLTVGSSAGLVAAASPVVAMSVTAGQAEITAAQVRAAAAYE 240
 DB 181 FSAASAFQSVVWGLTVGSSAGLVAAASPVVAMSVTAGQAEITAAQVRAAAAYE 240
 QY 241 TAYGLTVPPPPVIAENRAELMILITATNLGONTFAIAVNEAEYGEWMAQDAAMFGVAAAT 300
 DB 241 TAYGLTVPPPPVIAENRAELMILITATNLGONTFAIAVNEAEYGEWMAQDAAMFGVAAAT 300
 QY 301 ATATATLLPFEPAEMTSAGGLLEQAAAEEASDTAAANQLMNNVPOALQOLAQPTGGTT 360
 DB 301 ATATATLLPFEPAEMTSAGGLLEQAAAEEASDTAAANQLMNNVPOALQOLAQPTGGTT 360
 QY 361 PSSKLGGLWKTVPSPHRSPI SNMYSMANHMSMTNSGYVSMNTLSSMLKGFAPAAAOAVQ 420
 DB 361 PSSKLGGLWKTVPSPHRSPI SNMYSMANHMSMTNSGYVSMNTLSSMLKGFAPAAAOAVQ 420
 QY 421 TAAQNGVRAMSSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAAANQAVTPAARALP 480
 DB 421 TAAQNGVRAMSSSLGSSGLGGVAAANLGRAASVGSLSVPOAWAAANQAVTPAARALP 480
 QY 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSVLRVPPRPPVYVPHSPAGDIAPALS 540
 DB 481 LTSLSAAERGPQMLGGLPVQMGARAGGGLSVLRVPPRPPVYVPHSPAGDIAPALS 540
 QY 541 QDRPADFPALPLDPSAMVAQVPOVNAVNTKLGYNNAVAGTGTVIDPNGVLTNNHVA 600
 DB 541 QDRPADFPALPLDPSAMVAQVPOVNAVNTKLGYNNAVAGTGTVIDPNGVLTNNHVA 600
 QY 601 GATDINAFSVGGQTYGVYVVDYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGPPVAMGN 660
 DB 601 GATDINAFSVGGQTYGVYVVDYDRTQDVAVLQIRGAGGLPSAAIGGGVAVGPPVAMGN 660
 QY 661 SGGGGTTPRAVPRGVAVGQTVQASDSLTGAETLNGLIQFDAAIOPGSGGPPVNGLGQ 720
 DB 661 SGGGGTTPRAVPRGVAVGQTVQASDSLTGAETLNGLIQFDAAIOPGSGGPPVNGLGQ 720
 QY 721 VVGNTAAS 729
 DB 721 VVGNTAAS 729

RESULT 5
 ADA26366
 ID ADA26366 standard; protein; 825 AA.
 AC ADA26366;
 DT 20-NOV-2003 (first entry)
 DE Mycobacterium MTB72F-MTI (fusion MTB83F) protein.
 KW fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 KW tuberculosis; tuberculostatic; gene therapy; vaccine.
 OS Chimeric.
 OS Mycobacterium sp.
 PN W02003070187-A2.
 XX W02003070187-A2.
 PD 28-AUG-2003.
 PF 18-FEB-2003; 2003WO-US004903.
 PR 15-FEB-2002; 2002US-0357351P.
 XX (CORI-) CORIXA CORP.

PI Skeiky Y, Guderian J, Reed S;
XX MPI: 2003-697554/66.
DR N-PSDB; ADA26359.
XX
PT New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT and MTB39 antigens, with or without the MTB85A antigen, from a
PT Mycobacterium species, useful for diagnosing, preventing and/or treating
PT tuberculosis infection.
XX
PS Claim 85; Fig 14; 112pp; English.
XX
CC The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The
CC present sequence is used in the exemplification of the invention.
XX
SQ Sequence 825 AA;
Query Match 99.8%; Score 3680; DB 7; Length 825;
Best Local Similarity 99.9%; Pred. No. 3.7e-233;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MHHHHHTAASDNFOLSGGQGFAPIGQAMAIAGQIRSGGSPVHIGTAFGLGVVD 60
DB 1 MHHHHHTAASDNFOLSGGQGFAPIGQAMAIAGQIRSGGSPVHIGTAFGLGVVD 60
QY 61 NNGNGARVQVVGSAAPASLGISTGDDVTAVDGPINSATAMADALNGHHPGDVSVTWQ 120
DB 61 NNGNGARVQVVGSAAPASLGISTGDDVTAVDGPINSATAMADALNGHHPGDVSVTWQ 120
QY 121 TKSGETRTGNVTLAGPPAEFVDFGALPPEINARMYAGFGSASLVAAQAQWDSVASDL 180
DB 121 TKSGETRTGNVTLAGPPAEFVDFGALPPEINARMYAGFGSASLVAAQAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGWSIGSAGLVAAASPPYVAMSVTAGQAEALTAQAQVRAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGWSIGSAGLVAAASPPYVAMSVTAGQAEALTAQAQVRAAAAYE 240
QY 241 TAYGLTVPPVTAENRAELMILLIATNLGONTPTAVNAEYGENWAOADAAMFYAAT 300
DB 241 TAYGLTVPPVTAENRAELMILLIATNLGONTPTAVNAEYGENWAOADAAMFYAAT 300
QY 301 ATATATLTPFEAPEMTSAGGLEQAAVEEASDTAAANQLMNNVPQALQQLAQTQGT 360
DB 301 ATATATLTPFEAPEMTSAGGLEQAAVEEASDTAAANQLMNNVPQALQQLAQTQGT 360
QY 361 PSSKLGLLWKTVPSPHRSPISNMVMANNHMTNSGVMTNTLSSMLKGFAPAAARQAVQ 420
DB 361 PSSKLGLLWKTVPSPHRSPISNMVMANNHMTNSGVMTNTLSSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSLGSLGSGGVAANLGRAASVGLSVPOAWAANQAVTPAARALP 480
DB 421 TAAQNGVRAMSLGSLGSGGVAANLGRAASVGLSVPOAWAANQAVTPAARALP 480
QY 481 LTSLSAERPGQMLGGLPVQCMGARAGGGLSGVLVPPRPYVMPHSPAAGDIAPPALS 540
DB 481 LTSLSAERPGQMLGGLPVQCMGARAGGGLSGVLVPPRPYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVPOVNNINIKLYNNVAGAGTIVDPNGVLTNNHVIA 600
DB 541 QDRFADFPALPLDPSAMVAQVPOVNNINIKLYNNVAGAGTIVDPNGVLTNNHVIA 600
QY 601 GATDINAFVSGGQTVGVNDVYDRTQDVAVILQLRAGAGLPAAIIGGCVAVGEPPVAMGN 660
DB 601 GATDINAFVSGGQTVGVNDVYDRTQDVAVILQLRAGAGLPAAIIGGCVAVGEPPVAMGN 660
QY 661 SGGQGTTPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQDGGSGFPVNLGLQ 720

DB 661 SGGQGTTPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQDGGSGFPVNLGLQ 720
QY 721 VVGWNTAAS 729
DB 721 VVGWNTAAS 729
RESULT 6
ADA26365
ID ADA26365 standard; protein; 875 AA.
XX ADA26365;
XX
XX 20-NOV-2003 (first entry)
XX Mycobacterium MTB72F-Erd14 (fusion MTB89F) protein.
XX fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
XX tuberculosis; tuberculostatic; gene therapy; vaccine.
XX Chimeric.
XX Mycobacterium sp.
XX WO2003070187-A2.
XX
XX 28-AUG-2003.
XX
XX 18-FEB-2003; 2003WO-US004903.
XX
XX 15-FEB-2002; 2002US-0357351P.
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Guderian J, Reed S;
XX
XX MPI: 2003-697554/66.
XX N-PSDB; ADA26359.
XX
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
XX and MTB39 antigens, with or without the MTB85A antigen, from a
XX Mycobacterium species, useful for diagnosing, preventing and/or treating
XX tuberculosis infection.
XX
XX Claim 85; Fig 13; 112pp; English.
XX
CC The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The
CC present sequence is used in the exemplification of the invention.
XX
SQ Sequence 875 AA;
Query Match 99.8%; Score 3680; DB 7; Length 875;
Best Local Similarity 99.9%; Pred. No. 4e-233;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MHHHHHTAASDNFOLSGGQGFAPIGQAMAIAGQIRSGGSPVHIGTAFGLGVVD 60
DB 1 MHHHHHTAASDNFOLSGGQGFAPIGQAMAIAGQIRSGGSPVHIGTAFGLGVVD 60
QY 61 NNGNGARVQVVGSAAPASLGISTGDDVTAVDGPINSATAMADALNGHHPGDVSVTWQ 120
DB 61 NNGNGARVQVVGSAAPASLGISTGDDVTAVDGPINSATAMADALNGHHPGDVSVTWQ 120
QY 121 TKSGETRTGNVTLAGPPAEFVDFGALPPEINARMYAGFGSASLVAAQAQWDSVASDL 180
DB 121 TKSGETRTGNVTLAGPPAEFVDFGALPPEINARMYAGFGSASLVAAQAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGWSIGSAGLVAAASPPYVAMSVTAGQAEALTAQAQVRAAAAYE 240

Db 181 FSAASAFQSVVWGLTVGSSAGLWVAASAPYVWMSVTAQAEITAAQVRVAAAAYE 240
 Qy 241 TAYGLTVPPVPIAENRAELMILITATNLGONTFAIAVNEAEYGEKMAQDAAMFGVAAAT 300
 Db 241 TAYGLTVPPVPIAENRAELMILITATNLGONTFAIAVNEAEYGEKMAQDAAMFGVAAAT 300
 Qy 301 ATATATLLPPEEAPMTSAGLLEQAAVEASDTAAANQLMNVPAALQQAQPTQGT 360
 Db 301 ATATATLLPPEEAPMTSAGLLEQAAVEASDTAAANQLMNVPAALQQAQPTQGT 360
 Qy 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
 Db 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
 Qy 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOAWAAANQVTPAARALP 480
 Db 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOAWAAANQVTPAARALP 480
 Qy 481 LTSLSAAERGPQMLGGLPVGQMGARAGGLSGVLRVPPRYPMPSHSPAAGDIAPPALS 540
 Db 481 LTSLSAAERGPQMLGGLPVGQMGARAGGLSGVLRVPPRYPMPSHSPAAGDIAPPALS 540
 Qy 541 QDRFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVAGTGVIVDPNGVLTNNHVA 600
 Db 541 QDRFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVAGTGVIVDPNGVLTNNHVA 600
 Qy 601 GATDINAFSVGSGQTVGVVVDYRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMGN 660
 Db 601 GATDINAFSVGSGQTVGVVVDYRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMGN 660
 Qy 661 SGCGGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLQ 720
 Db 661 SGCGGTTPRAVGRVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVNGLQ 720
 Qy 721 VVGNTAAS 729
 Db 721 VVGNTAAS 729

RESULT 7
 AAE29731
 ID AAE29731 standard; protein; 930 AA.

XX AC AAE29731;
 XX DT 29-AUG-2003 (revised)
 XX DT 27-JAN-2003 (first entry)
 XX DE Mycobacterium sp. MTB72F-Leishmania sp. MAPS (aka r95f) fusion protein.

XX KW Vaccine; immunity; diagnostic agent; gene therapy; MTB72F; MAPS;
 XX fusion protein.

XX OS Mycobacterium sp.
 XX OS Leishmania sp.
 XX OS Chimeric.

XX PN W0200272792-A2.

XX PD 19-SEP-2002.

XX PF 13-MAR-2002; 2002WO-US008223.

XX PR 13-MAR-2001; 2001US-0275837P.

XX PA (CORI-) CORIXA CORP.

XX PI Skeiky Y, Brannon M, Guderian J;

XX DR WPI; 2002-759844/82.

XX DR N-PSDB; AAD47110.

PT New recombinant nucleic acid molecule comprising a Leishmania TSA, LeifF,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 PT tuberculosis.
 XX Example 6; Page 129-132; 155pp; English.

XX The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, LeifF, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is
 CC Mycobacterium sp. MTB72F-Leishmania sp. thiol specific antioxidant [TSA;
 CC MAPS (aka r95f)] fusion protein. This sequence comprises Mycobacterium
 CC species MTB72F (a 72 kDa poly-protein fusion construct comprising Rail2-
 CC Tbh9-Ra35) linked to the Leishmania TSA (MAPS). (Updated on 29-AUG-2003
 CC to standardise OS field)

XX Sequence 930 AA;

Query Match 99.8%; Score 3680; DB 5; Length 930;
 Best Local Similarity 99.9%; Pred. No. 4.3e-233;
 Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MHHHHHTAASDNFQLSQGGGGAIPGQAWATAGQIRSGGSGPTVHIGTAFGLGVVD 60
 Db 1 MHHHHHTAASDNFQLSQGGGGAIPGQAWATAGQIRSGGSGPTVHIGTAFGLGVVD 60
 Qy 61 NNGGARVQRVVGSAASLIGISTGVDITAVDGAIPNSATAMADALNGHHHPGDIVSVTWQ 120
 Db 61 NNGGARVQRVVGSAASLIGISTGVDITAVDGAIPNSATAMADALNGHHHPGDIVSVTWQ 120
 Qy 121 TKSGGTGTGNTLAEGPPAEFWDVFGALPPEINSARMYAGPGSASLVAAQMDVSASDL 180
 Db 121 TKSGGTGTGNTLAEGPPAEFWDVFGALPPEINSARMYAGPGSASLVAAQMDVSASDL 180
 Qy 181 FSAASAFQSVVWGLTVGSSAGLWVAASAPYVWMSVTAQAEITAAQVRVAAAAYE 240
 Db 181 FSAASAFQSVVWGLTVGSSAGLWVAASAPYVWMSVTAQAEITAAQVRVAAAAYE 240
 Qy 241 TAYGLTVPPVPIAENRAELMILITATNLGONTFAIAVNEAEYGEKMAQDAAMFGVAAAT 300
 Db 241 TAYGLTVPPVPIAENRAELMILITATNLGONTFAIAVNEAEYGEKMAQDAAMFGVAAAT 300
 Qy 301 ATATATLLPPEEAPMTSAGLLEQAAVEASDTAAANQLMNVPAALQQAQPTQGT 360
 Db 301 ATATATLLPPEEAPMTSAGLLEQAAVEASDTAAANQLMNVPAALQQAQPTQGT 360
 Qy 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
 Db 361 PSSKLGGLWKTVPSPHRSPISNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAVQ 420
 Qy 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOAWAAANQVTPAARALP 480
 Db 421 TAAQNGVRAMSSLGSSLGSGGVAANLGRAASVGSLSVPOAWAAANQVTPAARALP 480
 Qy 481 LTSLSAAERGPQMLGGLPVGQMGARAGGLSGVLRVPPRYPMPSHSPAAGDIAPPALS 540
 Db 481 LTSLSAAERGPQMLGGLPVGQMGARAGGLSGVLRVPPRYPMPSHSPAAGDIAPPALS 540
 Qy 541 QDRFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVAGTGVIVDPNGVLTNNHVA 600
 Db 541 QDRFADFPALPLDPSAMVAQVGPQVNVNINTKLGYNNAVAGTGVIVDPNGVLTNNHVA 600
 Qy 601 GATDINAFSVGSGQTVGVVVDYRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMGN 660

Db 601 GATDINAFSGGQTYGVVDVYDRTQDVAVLQRGAGGLPSAAIGGVAVGEPVVMAGN 660
 QY 661 SGGGGTPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQPGSGGPPVUNGLO 720
 Db 661 SGGGGTPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQPGSGGPPVUNGLO 720
 QY 721 VVGNTAAS 729
 Db 721 VVGNTAAS 729

RESULT 8
 ADA26364
 ID ADA26364 standard; protein; 930 AA.
 XX
 AC ADA26364;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Mycobacterium MTB72F-MAPS (fusion r95F) protein.
 XX
 KW fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 KW tuberculosis; tuberculostatic; gene therapy; vaccine.
 XX
 OS Chimeric.
 OS Mycobacterium sp.
 XX
 FN WO2003070187-A2.
 XX
 PD 28-AUG-2003.
 XX
 PF 18-FEB-2003; 2003WO-US004903.
 XX
 PR 15-FEB-2002; 2002US-0357351P.
 XX
 PA (CORI-) CORIXA CORP.
 PI Skeiky Y, Guderian J, Reed S;
 XX
 DR WPI; 2003-697554/66.
 XX
 DR N-PSDB; ADA26357.
 XX
 PT New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
 PT and MTB39 antigens, with or without the MTB85A antigen, from a
 PT Mycobacterium species, useful for diagnosing, preventing and/or treating
 PT tuberculosis infection.
 XX
 PS Disclosure; Fig 12; 112pp; English.
 XX

The invention relates to a novel nucleic acid encoding a fusion
 CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
 CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
 CC A polypeptide of the invention has tuberculostatic activity. A
 CC polynucleotide of the invention may have a use in gene therapy, and as a
 CC vaccine. The methods and compositions of the present invention are useful
 CC for diagnosing, preventing and/or treating tuberculosis infection. The
 CC present sequence is used in the exemplification of the invention.
 XX
 SQ Sequence 930 AA;

Query Match 99.8%; Score 3680; DB 7; Length 930;
 Best Local Similarity 99.9%; Pred. No. 4.3e-233;
 Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MEHHHHHTAASDNFOLSQGGQGFALPIGQMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
 Db 1 MEHHHHHTAASDNFOLSQGGQGFALPIGQMAIAGQIRSGGSPVHIGPTAFGLGVVD 60
 QY 61 NNGGARVORVVGSAAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
 Db 61 NNGGARVORVVGSAAPASLSIGTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
 QY 121 TKSGGTRIGNVTLAEGPFAEFWDFGALPPEINSARMYAGFGSASLVAAAQWDSVASDL 180

Db 121 TKSGGTRIGNVTLAEGPFAEFWDFGALPPEINSARMYAGFGSASLVAAAQWDSVASDL 180
 QY 181 FSAASAFOSVVGTLTVGSIWIGSSAGLMVAASAPYVAMSVTAGQAEULTAAQVRVAAAAAYE 240
 Db 181 FSAASAFOSVVGTLTVGSIWIGSSAGLMVAASAPYVAMSVTAGQAEULTAAQVRVAAAAAYE 240
 QY 241 TAYGLTVPPPIAENRAELMILIAITNLIGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 Db 241 TAYGLTVPPPIAENRAELMILIAITNLIGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 QY 301 ATATATLLPFEEAPEMTSAGLLEQAAAEEASPTAAANOLMNNVPQALQQLAQPTQTT 360
 Db 301 ATATATLLPFEEAPEMTSAGLLEQAAAEEASPTAAANOLMNNVPQALQQLAQPTQTT 360
 QY 361 PSSKLGLLWKTVPSPHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPALAAQAVQ 420
 Db 361 PSSKLGLLWKTVPSPHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPALAAQAVQ 420
 QY 421 TAAQNGVRAMSSLSGSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTTPAARALP 480
 Db 421 TAAQNGVRAMSSLSGSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTTPAARALP 480
 QY 481 LTSLSAAERPGOMLGLPVQOMGARAGGGLSGVLVPPRPYPMHPSPAAGDIAPPALS 540
 Db 481 LTSLSAAERPGOMLGLPVQOMGARAGGGLSGVLVPPRPYPMHPSPAAGDIAPPALS 540
 QY 541 QDRFADFPALPLDPSAMVAQVGPQVNVNITKLGYNNAVAGAGTGIVDPNGVLTNNHVA 600
 Db 541 QDRFADFPALPLDPSAMVAQVGPQVNVNITKLGYNNAVAGAGTGIVDPNGVLTNNHVA 600
 QY 601 GATDINAFSGGQTYGVVDVYDRTQDVAVLQRGAGGLPSAAIGGVAVGEPVVMAGN 660
 Db 601 GATDINAFSGGQTYGVVDVYDRTQDVAVLQRGAGGLPSAAIGGVAVGEPVVMAGN 660
 QY 661 SGGGGTPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQPGSGGPPVUNGLO 720
 Db 661 SGGGGTPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQPGSGGPPVUNGLO 720
 QY 721 VVGNTAAS 729
 Db 721 VVGNTAAS 729

RESULT 9
 ADA26370
 ID ADA26370 standard; protein; 1016 AA.
 XX
 AC ADA26370;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE M. bovis MTB72F and 85b complex antigen (fusion MTB103F).
 XX
 KW fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 KW tuberculosis; tuberculostatic; gene therapy; vaccine; MTB72F; 85B.
 XX
 OS Chimeric.
 OS Mycobacterium bovis.
 XX
 FN WO2003070187-A2.
 XX
 PD 28-AUG-2003.
 XX
 PF 18-FEB-2003; 2003WO-US004903.
 XX
 PR 15-FEB-2002; 2002US-0357351P.
 XX
 PA (CORI-) CORIXA CORP.
 PI Skeiky Y, Guderian J, Reed S;
 XX
 DR WPI; 2003-697554/66.

DR N-PSDB; ADA26363.
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT and MTB39 antigens, with or without the MTB85A antigen, from a
PT Mycobacterium species, useful for diagnosing, preventing and/or treating
PT tuberculosis infection.
XX
PS Claim 85; Fig 18; 112pp; English.
XX
XX The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The
CC present sequence is used in the exemplification of the invention.
XX
XX Sequence 1016 AA;
SQ
Query Match 99.8%; Score 3680; DB 7; Length 1016;
Best Local Similarity 99.9%; Pred. No. 4.8e-233;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MHHHHHTAASDNFQSGGGGFAIPGQAWAATAGQIRSGGGSPVHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQSGGGGFAIPGQAWAATAGQIRSGGGSPVHIGPTAFGLGVVD 60
QY 61 NNGNGARVORVVGSAAPASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
DB 61 NNGNGARVORVVGSAAPASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
QY 121 TKSQGTTRTGNVTLAEGPPAFPMVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
DB 121 TKSQGTTRTGNVTLAEGPPAFPMVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWSSAGLMVAASPVVAMSVTAGQAEITAAQVRAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSGWSSAGLMVAASPVVAMSVTAGQAEITAAQVRAAAAYE 240
QY 241 TAYGLTVPPPPVIAENRAELMILIAINLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
DB 241 TAYGLTVPPPPVIAENRAELMILIAINLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
QY 301 ATATATLLPFEAPEMTSAGLLLEQAAAEVEASDTAAANOLMNNVPOALQLOAQTOGTT 360
DB 301 ATATATLLPFEAPEMTSAGLLLEQAAAEVEASDTAAANOLMNNVPOALQLOAQTOGTT 360
QY 361 PSSKLGGLWKTQVSPHRSPISNNVSWANNHSMNTSGVSMNTLSSMLKGFAPAAQAQVQ 420
DB 361 PSSKLGGLWKTQVSPHRSPISNNVSWANNHSMNTSGVSMNTLSSMLKGFAPAAQAQVQ 420
QY 421 TAAQNGVRAMSSLGSSGLGGVGAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSGLGGVGAANLGRAASVGSLSVPQAWAANQAVTPAARALP 480
QY 481 LTSLSAERPGQMLGGLPVGQMGARAGGLSGVLVPPRPVYMPHSPAAAGDIAPALLS 540
DB 481 LTSLSAERPGQMLGGLPVGQMGARAGGLSGVLVPPRPVYMPHSPAAAGDIAPALLS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVYNINTKLGYNNAVAGAGTGVIPDPNGVLLTNHIVIA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPQVYNINTKLGYNNAVAGAGTGVIPDPNGVLLTNHIVIA 600
QY 601 GATDINAFSVSGQTYGVYDRTQDAVQLRGAGGLPSAAGGVAVGEPVWAMGN 660
DB 601 GATDINAFSVSGQTYGVYDRTQDAVQLRGAGGLPSAAGGVAVGEPVWAMGN 660
QY 661 SGGQCGTTPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQPGDSGGPVNGLGQ 720
DB 661 SGGQCGTTPRAVGRVVALGQTVQASDLSLTGAETLNGLIQFDAAIQPGDSGGPVNGLGQ 720
QY 721 VVGWMTAAS 729

DB 721 VVGWMTAAS 729
RESULT 10
ADA26369
ID ADA26369 standard; protein; 1022 AA.
XX
XX ADA26369;
AC
XX 20-NOV-2003 (first entry)
DT
XX Mycobacterium MTB72F-hrCC#1 (fusion MTB102Em2F) protein.
DE
XX fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW tuberculosis; tuberculostatic; gene therapy; vaccine.
XX
XX Chimeric.
OS
XX Mycobacterium sp.
XX
XX WO2003070187-A2.
FN
XX 28-AUG-2003.
PD
XX 18-FEB-2003; 2003WO-US004903.
PF
XX 15-FEB-2002; 2002US-0357351P.
PR
XX (CORI-) CORIXA CORP.
PA
XX Skeiky Y, Guderian J, Reed S;
PI
XX WPI: 2003-697554/66.
DR
XX N-PSDB; ADA26362.
XX
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT and MTB39 antigens, with or without the MTB85A antigen, from a
PT Mycobacterium species, useful for diagnosing, preventing and/or treating
PT tuberculosis infection.
XX
XX Claim 85; Fig 17; 112pp; English.
PS
XX The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The
CC present sequence is used in the exemplification of the invention.
XX
XX Sequence 1022 AA;
SQ
Query Match 99.8%; Score 3680; DB 7; Length 1022;
Best Local Similarity 99.9%; Pred. No. 4.8e-233;
Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MHHHHHTAASDNFQSGGGGFAIPGQAWAATAGQIRSGGGSPVHIGPTAFGLGVVD 60
DB 1 MHHHHHTAASDNFQSGGGGFAIPGQAWAATAGQIRSGGGSPVHIGPTAFGLGVVD 60
QY 61 NNGNGARVORVVGSAAPASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
DB 61 NNGNGARVORVVGSAAPASLGISTGDVITAVDGPINSATAMADALNGHHPGDVISTWQ 120
QY 121 TKSQGTTRTGNVTLAEGPPAFPMVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
DB 121 TKSQGTTRTGNVTLAEGPPAFPMVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
QY 181 FSAASAFQSVVWGLTVGSGWSSAGLMVAASPVVAMSVTAGQAEITAAQVRAAAAYE 240
DB 181 FSAASAFQSVVWGLTVGSGWSSAGLMVAASPVVAMSVTAGQAEITAAQVRAAAAYE 240

QY 241 TAYGLTVPVPIAENRAELMILITATLLGONTPTAVNEAEYGEWMAODAAAFYAAAT 300
DB 241 TAYGLTVPVPIAENRAELMILITATLLGONTPTAVNEAEYGEWMAODAAAFYAAAT 300
QY 301 ATATATLLPFEBAPENTSAGGLEQAAVEEASDTPAAANQLMNNVPOALQQAOTQGT 360
DB 301 ATATATLLPFEBAPENTSAGGLEQAAVEEASDTPAAANQLMNNVPOALQQAOTQGT 360
QY 361 PSSKLGGLWKTVPSPHRSPISNMVSNNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
DB 361 PSSKLGGLWKTVPSPHRSPISNMVSNNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSSLGSSGLGGGVAANLGRAASVGLSVPOAWAANQAVTPAARALP 480
DB 421 TAAQNGVRAMSSLGSSGLGGGVAANLGRAASVGLSVPOAWAANQAVTPAARALP 480
QY 481 LTSLSAAERGPQOMLGGLPVQMGARAGGGLSGVLVPPRPYVPHSPAAAGDIAPPALS 540
DB 481 LTSLSAAERGPQOMLGGLPVQMGARAGGGLSGVLVPPRPYVPHSPAAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVVLTNHVA 600
DB 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVVLTNHVA 600
QY 601 GATDINAFSVGSGQTYGVDRYDQDVAVLQLRGAGGLPSAAITGGVAVGEPVVMGN 660
DB 601 GATDINAFSVGSGQTYGVDRYDQDVAVLQLRGAGGLPSAAITGGVAVGEPVVMGN 660
QY 661 SGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQFGDSGGPVVNGLQ 720
DB 661 SGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQFGDSGGPVVNGLQ 720
QY 721 VVGNTAAS 729
DB 721 VVGNTAAS 729

RESULT 11

ADA26368

ID ADA26368 standard; protein; 1154 AA.

XX AC ADA26368;

XX DT 20-NOV-2003 (first entry)

XX DE Mycobacterium MTB72F-mTCC#2 (fusion MTB114F) protein.

XX KW fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;

XX KW tuberculosis; tuberculostatic; gene therapy; vaccine.

XX OS Chimeric.

XX OS Mycobacterium sp.

XX PN WO2003070187-A2.

XX PD 28-AUG-2003.

XX PF 18-FEB-2003; 2003WO-US004903.

XX PR 15-FEB-2002; 2002US-0357351P.

XX PA (CORI-) CORIXA CORP.

XX PI Skeiky Y, Guderian J, Reed S;

XX DR WPI; 2003-697554/66.

XX DR N-PSDB; ADA26361.

XX PT New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
XX PT and MTB39 antigens, with or without the MTB85A antigen, from a
XX PT Mycobacterium species, useful for diagnosing, preventing and/or treating
XX PT tuberculosis infection.

PS Claim 85; Fig 16; 112pp; English.

XX The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The
CC present sequence is used in the exemplification of the invention.

XX Sequence 1154 AA;

Query Match 99.8%; Score 3680; DB 7; Length 1154;

Best Local Similarity 99.9%; Pred. No. 5.6e-233;

Matches 728; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MEHHHTAASDNDNFQSQGGGFAIPGQAWATAGQIRSGGSPTVHIGTAFGLGVVD 60

DB 1 MEHHHTAASDNDNFQSQGGGFAIPGQAWATAGQIRSGGSPTVHIGTAFGLGVVD 60

QY 61 NNGNGARVQVRVGSAPAAASLGISTGVDVITAVDGAPINSATAMADALNHHHPGVDVISTWQ 120

DB 61 NNGNGARVQVRVGSAPAAASLGISTGVDVITAVDGAPINSATAMADALNHHHPGVDVISTWQ 120

QY 121 TKSGGTGTGNTVLAEGPPAFPMVDFGALPPEINSARMYAGPGSASLVAAQMDSVASDL 180

DB 121 TKSGGTGTGNTVLAEGPPAFPMVDFGALPPEINSARMYAGPGSASLVAAQMDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSGWISAGLMVAAAAPVVAAMSVTAGQAEITAAQVRAAAAYE 240

DB 181 FSAASAFQSVVWGLTVGSGWISAGLMVAAAAPVVAAMSVTAGQAEITAAQVRAAAAYE 240

QY 241 TAYGLTVPVPIAENRAELMILITATLLGONTPTAVNEAEYGEWMAODAAAFYAAAT 300

DB 241 TAYGLTVPVPIAENRAELMILITATLLGONTPTAVNEAEYGEWMAODAAAFYAAAT 300

QY 301 ATATATLLPFEBAPENTSAGGLEQAAVEEASDTPAAANQLMNNVPOALQQAOTQGT 360

DB 301 ATATATLLPFEBAPENTSAGGLEQAAVEEASDTPAAANQLMNNVPOALQQAOTQGT 360

QY 361 PSSKLGGLWKTVPSPHRSPISNMVSNNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420

DB 361 PSSKLGGLWKTVPSPHRSPISNMVSNNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420

QY 421 TAAQNGVRAMSSLGSSGLGGGVAANLGRAASVGLSVPOAWAANQAVTPAARALP 480

DB 421 TAAQNGVRAMSSLGSSGLGGGVAANLGRAASVGLSVPOAWAANQAVTPAARALP 480

QY 481 LTSLSAAERGPQOMLGGLPVQMGARAGGGLSGVLVPPRPYVPHSPAAAGDIAPPALS 540

DB 481 LTSLSAAERGPQOMLGGLPVQMGARAGGGLSGVLVPPRPYVPHSPAAAGDIAPPALS 540

QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVVLTNHVA 600

DB 541 QDRFADFPALPLDPSAMVAQVGPQVNNINTKLGYNNAVGAGTGIVDPNGVVLTNHVA 600

QY 601 GATDINAFSVGSGQTYGVDRYDQDVAVLQLRGAGGLPSAAITGGVAVGEPVVMGN 660

DB 601 GATDINAFSVGSGQTYGVDRYDQDVAVLQLRGAGGLPSAAITGGVAVGEPVVMGN 660

QY 661 SGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQFGDSGGPVVNGLQ 720

DB 661 SGGQGTTPRAVGRVVALGQTVQASDSLTAETLNGLIQFDAAIQFGDSGGPVVNGLQ 720

QY 721 VVGNTAAS 729

DB 721 VVGNTAAS 729

RESULT 12

AAE29709

ID AAE29709 standard; protein; 729 AA.

XX AC AAE29709;
 XX 29-AUG-2003 (revised)
 DT 27-JAN-2003 (first entry)
 XX
 XX Mycobacterium sp. MTB72FmutSA fusion protein.
 DE Vaccine; immunity; diagnostic agent; gene therapy; TbH9; antigen;
 KW Ra35MutSA; Ra12; MTB72MutSA; fusion protein.
 XX
 XX Mycobacterium sp.
 OS Mycobacterium tuberculosis.
 OS Chimeric.
 XX
 XX WO200272792-A2.
 FN 19-SEP-2002.
 PD
 XX 13-MAR-2002; 2002WO-US008223.
 PF
 XX 13-MAR-2001; 2001US-0275837P.
 PR (CORI-) CORIXA CORP.
 XX
 XX Skeiky Y, Brannon M, Guderian J;
 PI WPI; 2002-759844/82.
 XX N-ESDB; RAD47084.
 DR
 XX New recombinant nucleic acid molecule comprising a Leishmania TSA, Leif,
 PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
 PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
 PT tuberculosis.
 XX
 XX Disclosure; Page 93-95; 155pp; English.
 XX
 XX The invention relates to a recombinant nucleic acid molecule encoding a
 CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
 CC polynucleotide sequence encoding an antigen or an antigenic fragment from
 CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
 CC polypeptide or its fragment. The Leishmania polynucleotide is selected
 CC from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention
 CC are used in methods for eliciting immune response in mammals. They are
 CC useful as vaccines to elicit protective immunity against pathogenic
 CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
 CC polypeptides are used for enhancing the expression of polynucleotides, as
 CC in vivo diagnostic agents and for raising antibodies in a non-human
 CC animal. The invention is used in gene therapy. The present sequence is
 CC MTB72F fusion protein. This fusion protein comprises Ra35MutSA mutant
 CC protein and Ra12 protein from M. tuberculosis and TbH9 protein from
 CC Mycobacterium sp. (Updated on 29-AUG-2003 to standardise OS field)
 XX
 XX Sequence 729 AA;
 SQ
 Query Match 99.8%; Score 3677; DB 5; Length 729;
 Best Local Similarity 99.7%; Pred. No. 5e-233;
 Matches 727; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MHHHHHTAASDNFQLSQGGGPAIPGQAMATAGQIRSGGSPVTHIGTPFLGIVVD 60
 DB 1 MHHHHHTAASDNFQLSQGGGPAIPGQAMATAGQIRSGGSPVTHIGTPFLGIVVD 60
 QY 61 NNGNGARVQVWGSAPASLIGISTGVDITAVDGPINSATAMADALNGHPGDIVISVTWQ 120
 DB 61 NNGNGARVQVWGSAPASLIGISTGVDITAVDGPINSATAMADALNGHPGDIVISVTWQ 120
 QY 121 TKSGGTGTHVTLAEGPAPFMDVFGNLPPEINSARMYAGPGSASLVAAQWDSVASDL 180
 DB 121 TKSGGTGTHVTLAEGPAPFMDVFGNLPPEINSARMYAGPGSASLVAAQWDSVASDL 180
 QY 181 FSAASAFQSVVWGLTVGSWITGSSAGLMVAAASPYVAMNSVTAGQAEITAAQVRVAAAAYE 240

DB 181 FSAASAFQSVVWGLTVGSWITGSSAGLMVAAASPYVAMNSVTAGQAEITAAQVRVAAAAYE 240
 QY 241 TAYGLTVPPVIAENRAELMILITATNLIGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 DB 241 TAYGLTVPPVIAENRAELMILITATNLIGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 QY 301 ATATATLTPFEAEPEMTSAGLLQFAAAVEBASDTAAANQLMNNVPOALQOLAQPTOGTT 360
 DB 301 ATATATLTPFEAEPEMTSAGLLQFAAAVEBASDTAAANQLMNNVPOALQOLAQPTOGTT 360
 QY 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMTNSGVSMTNTLSSMLKGPAPAAAQAVQ 420
 DB 361 PSSKLGGLWKTVPSPHRSPISNMVMANNHMTNSGVSMTNTLSSMLKGPAPAAAQAVQ 420
 QY 421 TAAQNGVRAMSSLSGSSLSGSSGVAANLGRASVSGSLSVPOAWAANQAVTPAARALP 480
 DB 421 TAAQNGVRAMSSLSGSSLSGSSGVAANLGRASVSGSLSVPOAWAANQAVTPAARALP 480
 QY 481 LTSLSAERGPOMLGLPVGQMGARAGGSLGVLPVPPPYVMHSPHSPAGDTAPPALS 540
 DB 481 LTSLSAERGPOMLGLPVGQMGARAGGSLGVLPVPPPYVMHSPHSPAGDTAPPALS 540
 QY 541 QDRFADFPALPLDPSAMVAQVGPQVNVINTKLGYNNAVAGCTGIVDPNGVLLTNNHVA 600
 DB 541 QDRFADFPALPLDPSAMVAQVGPQVNVINTKLGYNNAVAGCTGIVDPNGVLLTNNHVA 600
 QY 601 GATDINAFSVGSGQTYGVDVVDYDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVWANG 660
 DB 601 GATDINAFSVGSGQTYGVDVVDYDRTQDVAVLQLRGAGLPSAAIGGGVAVGEPVWANG 660
 QY 661 SGGGGTTPRVPGRVVALGQTVQASDSLSITGAETLNGLIQFDAAIQPCDSGGPVVNGLGQ 720
 DB 661 SGGGGTTPRVPGRVVALGQTVQASDSLSITGAETLNGLIQFDAAIQPCDSGGPVVNGLGQ 720
 QY 721 VVGMMNTAAS 729
 DB 721 VVGMMNTAAS 729
 RESULT 13
 AAEI7573
 ID AAEI7573 standard; protein; 729 AA.
 XX
 AC AAEI7573;
 XX
 DT 22-APR-2002 (first entry)
 XX Mycobacterium species MTB72FmutSA fusion protein.
 DE Fusion protein; antigen; serological sensitivity; immune response;
 KW tuberculosis; infection; vaccine; MTB72FmutSA; Ra12-TbH9-Ra35MutSA;
 KW mutant; mutein.
 XX
 OS Mycobacterium sp.
 XX
 FH Key Location/Qualifiers
 FT Region 8..139
 FT Region /label= Ra12_protein
 FT Region 142..532
 FT Region /label= TbH9FL_protein
 FT Region 535..729
 FT Region /label= Ra35_protein
 FT Misc-difference 710
 FT /note= "Wild type Ser substituted with Ala"
 XX
 PN WO200198460-A2.
 XX
 PD 27-DEC-2001.
 XX
 XX 20-JUN-2001; 2001WO-US019959.
 XX
 XX 20-JUN-2000; 2000US-00597796.
 PR 01-FEB-2001; 2001US-0265737P.

XX (CORI-) CORIXA CORP.
XX Skeiky Y, Reed S, Alderson M;
XX WPI; 2002-147798/19.
XX N-PSDB; AAD28343.
XX Composition comprising MTB39 antigen and MTB32A antigen from
PT Mycobacterium species, useful for eliciting immune response in a subject.
XX Claim 88; Fig 5; 136pp; English.
XX The present invention relates to fusion proteins containing at least two
CC Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected with
CC Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
CC MTB72FmutSA (Ra12-TbH9-Ra35MutSA) mutant fusion protein
XX MTB72FmutSA (Ra12-TbH9-Ra35MutSA) mutant fusion protein
XX Sequence 729 AA;
Query Match 99.8%; Score 3677; DB 5; Length 729;
Best Local Similarity 99.7%; Pred. No. 5e-233;
Matches 727; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MEHHHHTAASDNFQSQGGQGFAPIGQAMAIAGIRSGGSPVHIGPTAFGLGVVD 60
Db 1 MEHHHHTAASDNFQSQGGQGFAPIGQAMAIAGIRSGGSPVHIGPTAFGLGVVD 60
QY 61 NNGNGARVQVVGSAAPASLGI-STGDVITAVDGPINSATAMADALNGHPGDVISTWQ 120
Db 61 NNGNGARVQVVGSAAPASLGI-STGDVITAVDGPINSATAMADALNGHPGDVISTWQ 120
QY 121 TKSGGTGTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDL 180
Db 121 TKSGGTGTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDL 180
QY 181 FSAASAFQSVWGLTVGSGWIGSAGLWMAAASPYVAMSVTAQAEITAAQVRAAAAYE 240
Db 181 FSAASAFQSVWGLTVGSGWIGSAGLWMAAASPYVAMSVTAQAEITAAQVRAAAAYE 240
QY 241 TAYGLTVPVPPVIAENRAELMILIAITMLLCQNTPAIAVNEAEYGEWQAQAAAMFGYAAAT 300
Db 241 TAYGLTVPVPPVIAENRAELMILIAITMLLCQNTPAIAVNEAEYGEWQAQAAAMFGYAAAT 300
QY 301 ATATATLLPPEAPEMTSAGGLLEQAAVEEASDTAAANCLMNNVPOALQQLAQPTGTT 360
Db 301 ATATATLLPPEAPEMTSAGGLLEQAAVEEASDTAAANCLMNNVPOALQQLAQPTGTT 360
QY 361 PSSKLGGLWKTVSPHRSPISSNMVSMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
Db 361 PSSKLGGLWKTVSPHRSPISSNMVSMANNHMTNSGVSMTNTLSSMLKGFAPAAARQAVQ 420
QY 421 TAAQNGVRAMSLGSSLSGSGGGVAAANLGRAASVGSLSVQAAVAAANQAVTPBARALP 480
Db 421 TAAQNGVRAMSLGSSLSGSGGGVAAANLGRAASVGSLSVQAAVAAANQAVTPBARALP 480

QY 481 LTSLSAAERGPQMLGGLPVGQMGARAGGGLSGVLRRPVPYVMPHSPAAGDIAPPALS 540
Db 481 LTSLSAAERGPQMLGGLPVGQMGARAGGGLSGVLRRPVPYVMPHSPAAGDIAPPALS 540
QY 541 QDRFADFPALPLDPSAMVAQVGVNINTKLGYNNAVAGCTGIVIDPNCVLTNNHVIA 600
Db 541 QDRFADFPALPLDPSAMVAQVGVNINTKLGYNNAVAGCTGIVIDPNCVLTNNHVIA 600
QY 601 GATDINAFSGSGQTYGVVDVVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGPPVAMGN 660
Db 601 GATDINAFSGSGQTYGVVDVVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGPPVAMGN 660
QY 661 SGGGGTTPRAVPRGRVVALGQTVQASDSLTGAETLNGLIQFDAAIOPGDSGGPVNGLGQ 720
Db 661 SGGGGTTPRAVPRGRVVALGQTVQASDSLTGAETLNGLIQFDAAIOPGDSGGPVNGLGQ 720
QY 721 VVGNNMTAAS 729
Db 721 VVGNNMTAAS 729
RESULT 14
ADA26374
ID ADA26374 standard; protein; 729 AA.
XX ADA26374;
AC ADA26374;
DT 20-NOV-2003 (first entry)
DE Mycobacterium MTB72FmutSA protein.
XX MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A; tuberculosis;
KW tuberculostatic; Gene therapy; vaccine; MTB72F; mutant; mutcin.
XX Synthetic.
OS Mycobacterium sp.
XX Key Location/Qualifiers
FT Misc-difference 710 /note= "Wild-type Ser substituted by Ala"
XX WO2003070187-A2.
PD 28-AUG-2003.
PF 18-FEB-2003; 2003WO-US004903.
XX 15-FEB-2002; 2002US-0357351P.
PA (CORI-) CORIXA CORP.
XX Skeiky Y, Guderian J, Reed S;
XX WPI; 2003-697554/66.
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT MTB39 antigens, with or without the MTB85A antigen, from a
PT Mycobacterium species, useful for diagnosing, preventing and/or treating
PT tuberculosis infection.
PS Disclosure; Fig 20; 112pp; English.
XX The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The
CC present sequence is used in the exemplification of the invention.
XX Sequence 729 AA;

Query Match 99.8%; Score 3677; DB 7; Length 729;
 Best Local Similarity 99.7%; Pred. No. 5e-233;
 Matches 727; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAWATAGQIRSGGSPVTHIGPTAFGLGVVD 60
 DB 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAWATAGQIRSGGSPVTHIGPTAFGLGVVD 60

QY 61 NNGNGARVQRVVGSAPASLIGSTGDVITAVDGAPINSATAMADALNGHHPGDVISVTWQ 120
 DB 61 NNGNGARVQRVVGSAPASLIGSTGDVITAVDGAPINSATAMADALNGHHPGDVISVTWQ 120

QY 121 TKSGETTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
 DB 121 TKSGETTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSWIGSSAGLVAAASPVVAMNSVTAGQAEITAAQVRAAAAYE 240
 DB 181 FSAASAFQSVVWGLTVGSWIGSSAGLVAAASPVVAMNSVTAGQAEITAAQVRAAAAYE 240

QY 241 TAYGLTVPPVPIAENRAELMILLIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 DB 241 TAYGLTVPPVPIAENRAELMILLIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300

QY 301 ATATATLLPPEEAPMTSAGGLLEQAAAVEASDSTAANQLMNNVPOALQOLAQPTGTT 360
 DB 301 ATATATLLPPEEAPMTSAGGLLEQAAAVEASDSTAANQLMNNVPOALQOLAQPTGTT 360

QY 361 PSSKLGGLWKTVSPHRSPISNMVSMMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
 DB 361 PSSKLGGLWKTVSPHRSPISNMVSMMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420

QY 421 TAAQNGVRAMSSLGSSLGSSGLGGVAAANLGRAASVCSLSVPQAWAANQAVTPAARALP 480
 DB 421 TAAQNGVRAMSSLGSSLGSSGLGGVAAANLGRAASVCSLSVPQAWAANQAVTPAARALP 480

QY 481 LTSLSAARPGQMLGGLPVGQMGARAGGLSGVLVRPPVPMPSHAAGDIAPPALS 540
 DB 481 LTSLSAARPGQMLGGLPVGQMGARAGGLSGVLVRPPVPMPSHAAGDIAPPALS 540

QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNIKLGYNNAVGAGTGIVDPNGVLVTHNHVIA 600
 DB 541 QDRFADFPALPLDPSAMVAQVGPQVNNIKLGYNNAVGAGTGIVDPNGVLVTHNHVIA 600

QY 601 GATDINAFSVSGQTYGVVVDVDRDQDAVQLRGAGGLPSAAIGGVAVGEPVVMAGN 660
 DB 601 GATDINAFSVSGQTYGVVVDVDRDQDAVQLRGAGGLPSAAIGGVAVGEPVVMAGN 660

QY 661 SGGGGTTPRAVPGRVVAGLQTVQASDSLTGAETLNLGLIOFDAIOPGDSGGPVVNLGQ 720
 DB 661 SGGGGTTPRAVPGRVVAGLQTVQASDSLTGAETLNLGLIOFDAIOPGDSGGPVVNLGQ 720

QY 721 VVGNTAAS 729
 DB 721 VVGNTAAS 729

RESULT 15
 ADA26373
 ID ADA26373 standard; protein; 729 AA.
 XX
 AC ADA26373;
 XX
 DT 20-NOV-2003 (first entry)
 XX
 DE Mycobacterium wild type MBT72F protein.
 XX
 KW fusion polypeptide; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
 XX tuberculosis; tuberculostatic; gene therapy; vaccine; MTB72F.
 OS Mycobacterium sp.
 XX
 PN W02003070187-A2.

XX 28-AUG-2003.
 PD 18-FEB-2003; 2003WO-US004903.
 XX 15-FEB-2002; 2002US-0357351P.
 PR (CORI-) CORIXA CORP.
 PA Skeiky Y, Guderian J, Reed S;
 PI WPI; 2003-697554/66.
 DR
 XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
 PT and MTB39 antigens, with or without the MTB85A antigen, from a
 PT Mycobacterium species, useful for diagnosing, preventing and/or treating
 PT tuberculosis infection.
 PS Disclosure; fig 20; 112pp; English.
 XX The invention relates to a novel nucleic acid encoding a fusion
 CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
 CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
 CC A polypeptide of the invention has tuberculostatic activity. A
 CC polynucleotide of the invention may have a use in gene therapy, and as a
 CC vaccine. The methods and compositions of the present invention are useful
 CC for diagnosing, preventing and/or treating tuberculosis infection. The
 CC present sequence is used in the exemplification of the invention.
 XX Sequence 729 AA;

Query Match 99.6%; Score 3670; DB 7; Length 729;
 Best Local Similarity 99.7%; Pred. No. 1.4e-232;
 Matches 727; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAWATAGQIRSGGSPVTHIGPTAFGLGVVD 60
 DB 1 MHHHHHTAASDNFQLSQGGGFAIPIGQAWATAGQIRSGGSPVTHIGPTAFGLGVVD 60

QY 61 NNGNGARVQRVVGSAPASLIGSTGDVITAVDGAPINSATAMADALNGHHPGDVISVTWQ 120
 DB 61 NNGNGARVQRVVGSAPASLIGSTGDVITAVDGAPINSATAMADALNGHHPGDVISVTWQ 120

QY 121 TKSGETTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180
 DB 121 TKSGETTGNVTLAEGPPAEFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDL 180

QY 181 FSAASAFQSVVWGLTVGSWIGSSAGLVAAASPVVAMNSVTAGQAEITAAQVRAAAAYE 240
 DB 181 FSAASAFQSVVWGLTVGSWIGSSAGLVAAASPVVAMNSVTAGQAEITAAQVRAAAAYE 240

QY 241 TAYGLTVPPVPIAENRAELMILLIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300
 DB 241 TAYGLTVPPVPIAENRAELMILLIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAAT 300

QY 301 ATATATLLPPEEAPMTSAGGLLEQAAAVEASDSTAANQLMNNVPOALQOLAQPTGTT 360
 DB 301 ATATATLLPPEEAPMTSAGGLLEQAAAVEASDSTAANQLMNNVPOALQOLAQPTGTT 360

QY 361 PSSKLGGLWKTVSPHRSPISNMVSMMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420
 DB 361 PSSKLGGLWKTVSPHRSPISNMVSMMANNHMTNSGVSMTNTLSSMLKGFAPAAAQAVQ 420

QY 421 TAAQNGVRAMSSLGSSLGSSGLGGVAAANLGRAASVCSLSVPQAWAANQAVTPAARALP 480
 DB 421 TAAQNGVRAMSSLGSSLGSSGLGGVAAANLGRAASVCSLSVPQAWAANQAVTPAARALP 480

QY 481 LTSLSAARPGQMLGGLPVGQMGARAGGLSGVLVRPPVPMPSHAAGDIAPPALS 540
 DB 481 LTSLSAARPGQMLGGLPVGQMGARAGGLSGVLVRPPVPMPSHAAGDIAPPALS 540

QY 541 QDRFADFPALPLDPSAMVAQVGPQVNNIKLGYNNAVGAGTGIVDPNGVLVTHNHVIA 600

Db 541 QDEFADFPALPLDPSAMVAQVGPQVYNINTKLGYNNNAVAGGTGIVIDPNGVVLNNHVIA 600
QY 601 GATDINAFVSGQTYGVDVWGYDRTQDVAVLQLRGAGGLPSAAICGGYAVGEPVWAMGN 660
Db 601 GATDINAFVSGQTYGVDVWGYDRTQDVAVLQLRGAGGLPSAAICGGYAVGEPVWAMGN 660
QY 661 SGGQGGTPRAVPGRVVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVWNLGQ 720
Db 661 SGGQGGTPRAVPGRVVVALGQTVQASDSLTCABETLNGLIQFDAAIQPGDSGGPVVWNLGQ 720
QY 721 VVGNTAAS 729
Db 721 VVGNTAAS 729

Search completed: June 30, 2004, 16:48:51
Job time : 76.9614 secs

THESE DOCUMENTS SONT
DEPOSES A LA
BIBLIOTHEQUE
NATIONALE
FRANCAISE
PAR
LE
SERVISE
NATIONAL
DE
L'INFORMATION
SCIENTIFIQUE
ET
TECHNIQUE
EN
FRA
NCE

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:01 ; Search time 43.4105 Seconds

(without alignments)
4331.879 Million cell updates/sec

Title: US-09-597-796C-10

Perfect score: 3007

Sequence: 1 HHHHHHHVDFGALPPEIN.....SGGPVVGGLGVGVGNTAAS 596

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1949	64.8	391	005298	005298 mycobacteri
2	1920.5	63.9	390	16 Q7U089	Q7U089 mycobacteri
3	1779.5	59.2	396	16 Q7U071	Q7U071 mycobacteri
4	1648	54.8	393	16 Q8V123	Q8V123 mycobacteri
5	1583	52.6	393	16 Q06341	Q06341 mycobacteri
6	1583	52.6	393	16 Q7TWFS	Q7TWFS mycobacteri
7	989	32.9	355	16 Q07175	Q07175 mycobacteri
8	989	32.9	355	16 Q7U289	Q7U289 mycobacteri
9	779.5	25.9	393	16 Q7TZJ3	Q7TZJ3 mycobacteri
10	779.5	25.9	411	16 Q53939	Q53939 mycobacteri
11	778.5	25.9	361	2 Q50320	Q50320 mycobacteri
12	756.5	25.2	410	2 Q99Q11	Q99Q11 mycobacteri
13	754.5	25.1	354	16 Q5CCY9	Q5CCY9 mycobacteri
14	753.5	25.1	409	16 Q53957	Q53957 mycobacteri
15	751.5	25.0	409	16 Q7TZH7	Q7TZH7 mycobacteri
16	740.5	24.6	399	16 Q7TZH8	Q7TZH8 mycobacteri

17	737.5	24.5	403	16	053956	053956 mycobacteri
18	737	24.5	423	16	053950	053950 mycobacteri
19	737	24.5	423	16	07TZ14	07TZ14 mycobacteri
20	731	24.3	421	16	Q9Z5K0	Q9Z5K0 mycobacteri
21	702	23.3	391	16	P96362	P96362 mycobacteri
22	702	23.3	391	16	Q7U075	Q7U075 mycobacteri
23	702	23.3	413	16	Q06386	Q06386 mycobacteri
24	695	23.1	694	16	Q8VJW0	Q8VJW0 mycobacteri
25	690	22.9	468	16	053958	053958 mycobacteri
26	675	22.4	380	16	P95190	P95190 mycobacteri
27	675	22.4	380	16	Q7TX66	Q7TX66 mycobacteri
28	670.5	22.3	385	16	Q7TZR7	Q7TZR7 mycobacteri
29	669.5	22.3	394	16	Q7TXX5	Q7TXX5 mycobacteri
30	668.5	22.2	462	16	Q33110	Q33110 mycobacteri
31	667.5	22.2	385	16	Q33204	Q33204 mycobacteri
32	666.5	22.2	385	16	Q8VJZ0	Q8VJZ0 mycobacteri
33	641	21.3	364	16	Q7TZJ5	Q7TZJ5 mycobacteri
34	637	21.2	350	16	Q7TZJ2	Q7TZJ2 mycobacteri
35	635	21.1	363	16	O53940	O53940 mycobacteri
36	633.5	21.1	365	16	O86373	O86373 mycobacteri
37	629.5	20.9	405	16	O8VJW5	O8VJW5 mycobacteri
38	626.5	20.8	381	16	Q7TX67	Q7TX67 mycobacteri
39	620.5	20.6	443	16	Q8VKL9	Q8VKL9 mycobacteri
40	618.5	20.6	397	2	Q9AGF0	Q9AGF0 mycobacteri
41	615.5	20.5	382	16	Q7TXX3	Q7TXX3 mycobacteri
42	615.5	20.5	402	16	O33312	O33312 mycobacteri
43	614.5	20.4	443	16	Q7U242	Q7U242 mycobacteri
44	606.5	20.2	423	16	Q7U114	Q7U114 mycobacteri
45	606.5	20.2	426	16	O05907	O05907 mycobacteri

ALIGNMENTS

RESULT 1
005298
ID O05298 PRELIMINARY; PRT; 391 AA.
AC O05298;
DT 01-JUL-1997 (TRENBLREL. 04, Created)
DT 01-JUL-1997 (TRENBLREL. 04, Last sequence update)
DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)
DE Hypothetical protein (PPE family protein).
GN RV1196 OR MTC1364.08 OR MTL234.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaiia F.,
Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
Hornsby T., Jagals K., Krogh A., McLean J., Moule S., Murphy L.,
Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
Kolony J.F., Nelson W.C., Umayam L.A., Emdolaeva M.D., Salzberg S.L.,
Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
Bisshai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
laboratory strains";
RT Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

Query Match 63.9%; Score 1920.5; DB 16; Length 390;
Best Local Similarity 99.0%; -Pred. No. 9.6e-89;
Matches 387; Conservative 2; Mismatches 1; Indels 1; Gaps 1;


```
Db 300 GSSGLGAGVAANLGRASVGLSVPPAWAANAQVTPAARALPLTLTSLTSAAGTAPGHMLG 359
Qy 365 GLPVGMGARAGGLGVLRLVPRPVVMPHSPAG 399
Db 360 GLPLGH-SVNAGSGINNLRVPAAYAIPTPTPAAG 393

RESULT 6
Q7TWFS
ID Q7TWFS PRELIMINARY; PRT; 393 AA.
AC Q7TWFS
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE PE family protein.
GN PPE60 OR MB3505.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AP2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Biglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duchoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.,
RA "The complete genome sequence of Mycobacterium bovis."
RA Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
RL EMBL: BX248346; CAD95692.1; -.
RW Complete proteome.
KW SEQUENCE 393 AA; 39413 MW; AF4C20C95DAB7DD4 CRC64;

Query Match 52.6%; Score 1583; DB 16; Length 393;
Best Local Similarity 81.5%; Pred. No. 7.8e-72;
Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

Qy 9 MVDFGALPEINSARMYAGSGSLVAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
Db 1 MVDFGALPEINSARMYAGSGSLVAQAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
Qy 69 SSAGLVAAASPVVWMSVTAGQAEITAAQVRVAAAYETAYCLTVPPVPIAENRAELMI 128
Db 61 SSAGLVAAASPVVWMSVTAGQAEITAAQVRVAAAYETAYCLTVPPVPIAENRAELMT 120
Qy 129 LIATNLLGQNTPAIAVNEAYEGEMWADAAAFGVAATATATATATATATATATATATAT 188
Db 121 LIATNLLGQNTPAIAVNEAYEGEMWADAAAFGVAATATATATATATATATATATATAT 180
Qy 189 LLEQAAVVEASDTAAANOLMNNVPALQALQAOPTOGTTTPSSKLGGLMKTVPSPHSPI 248
Db 181 LLEQAAVVEASDTAAANOLMNNVPALQALQAOPTOGTTTPSSKLGGLMKTVPSPHSPI 240
Qy 249 MVSMAHHMSMTSGVSMNTTSLMLKGPAPAAAQAVTAQNGVRAMS-----LGSSL 304
Db 241 VSSIANNHMSMTSGVSMNTTSLMLKGPAPAAAQAVTAQNGVRAMS-----LGSSL 299
Qy 305 GSSGLGGVAANLGRASVGLSVPCAANAQVTPAARALPLTLTSLTSAAGRGQMLG 364
Db 300 GSSGLGAGVAANLGRASVGLSVPCAANAQVTPAARALPLTLTSLTSAAGTAPGHMLG 359
Qy 365 GLPVGMGARAGGLGVLRLVPRPVVMPHSPAG 399
Db 360 GLPLGH-SVNAGSGINNLRVPAAYAIPTPTPAAG 393

RESULT 7
O07175
ID O07175 PRELIMINARY; PRT; 355 AA.
AC O07175
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
```

```
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein (Serine protease, putative).
GN PEPA OR RV0125 OR MTC1418B.07 OR MT0133.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RA complete genome sequence."
RA Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.D., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RA "Whole genome comparison of Mycobacterium tuberculosis clinical and
RA laboratory strains."
RA Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO SERINE PROTEASES, TRYPsin FAMILY.
DR EMBL: Z96071; CAB09453.1; -.
DR EMBL: AE006925; AAK44357.1; -.
DR PIR: F70983; F70983.
DR TIGR: MT0133; -.
DR Tuberculist: Rv0125; -.
DR GO: GO:0004295; P:trypsin activity; IEA.
DR GO: GO:0007242; P:intracellular signaling cascade; IEA.
DR GO: GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro: IPR009003; Cys_Ser_trypsin.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001254; Peptidase_S1.
DR InterPro: IPR008256; Peptidase_S1B_V8.
DR InterPro: IPR001940; Peptidase_S1C.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR PRINTS: PR00839; V8PROTEASE.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS00106; PDZ; 1.
DR PROSITE: PS00135; TRYPsin_SER; 1.
KW Hydrolase, Hypothetical protein, Serine protease, Protease;
KW Complete proteome.
SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match 32.9%; Score 989; DB 16; Length 355;
Best Local Similarity 98.0%; Pred. No. 3.9e-42;
Matches 196; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 397 AAGDIAPPALSQDRFADFPALPLDPSAMVAQGVQVNNINIKLGYNNVAGAGTIVIDPN 456
Db 28 APAQAAPPALSQDRFADFPALPLDPSAMVAQGVQVNNINIKLGYNNVAGAGTIVIDPN 87
Qy 457 GVVLITNNHVITAGATDINAFSVGSGQYGVVDVGVYDRTQDVAVLQLRGAGLPSAAIGGGV 516
Db 88 GVVLITNNHVITAGATDINAFSVGSGQYGVVDVGVYDRTQDVAVLQLRGAGLPSAAIGGGV 147
Qy 517 AVGPEVVAANGSGQGTQPRVPRVVALGQTVQASDLSLTGAETLNLGIFQDAAIOPGD 576
```



```

Db 148 AVGEPVVMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGD 207
QY 577 SGGPVVNGLGQVVGWNTAAS 596
Db 208 SGGPVVNGLGQVVGWNTAAS 227

RESULT 8
QYU2S9 PRELIMINARY; PRT; 355 AA.
AC QYU2S9;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DE Probable serine protease PEPA (EC 3.4.21.-).
GN PEPA OR MB0130.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eglmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
RL EMBL; BX248334; CAD92991.1; -.
KW Hydrolyase; Complete proteome.
SQ SEQUENCE 355 AA; 34926 MW; 16CE9E21A97BF192 CRC64;

Query Match 32.9%; Score 989; DB 16; Length 355;
Best Local Similarity 98.0%; Pred. No. 3,9e-42;
Matches 196; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 397 AAGDIAPALSDQRFADFPALPLDPSAMVAVQGVPOVNVNTKLGYNNAVGAGTGIVIDPN 456
Db 28 APAQAAPPALSDQRFADFPALPLDPSAMVAVQGVPOVNVNTKLGYNNAVGAGTGIVIDPN 87
QY 457 GVVLNNHVIAGATDINAFSGVSGQTYGVVDVDRTOAVLQIRGAGGLPSAIGGV 516
Db 88 GVVLNNHVIAGATDINAFSGVSGQTYGVVDVDRTOAVLQIRGAGGLPSAIGGV 147
QY 517 AVGEPVVMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGD 576
Db 148 AVGEPVVMGNSGGGGTTPRAVGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPGD 207
QY 577 SGGPVVNGLGQVVGWNTAAS 596
Db 208 SGGPVVNGLGQVVGWNTAAS 227

RESULT 9
QYU2J3 PRELIMINARY; PRT; 393 AA.
AC QYU2J3;
DT 01-OCT-2003 (TReMBLrel. 25, Created)
DT 01-OCT-2003 (TReMBLrel. 25, Last sequence update)
DE PPE family protein.
GN PPE26 OR MB1817.
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1765;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eglmeier K., Camus J.-C., Medina N., Mansoor H.,

```

```

RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
RL EMBL; BX248340; CAD94520.1; -.
KW Complete proteome.
SQ SEQUENCE 393 AA; 38572 MW; 41944B6E547A8AB0 CRC64;

Query Match 25.9%; Score 779.5; DB 16; Length 393;
Best Local Similarity 43.7%; Pred. No. 1.4e-31;
Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;

QY 10 VDFGALPPEINSAVMYAGPGSASLVAAQWDSVASDLSAASAFQSVWGLTVGWSIGS 69
Db 1 MDFGALPPEINSAVMYAGPGSAPMWAAASAWNGLAELSAAATGYETVITQLSSEGLGP 60
QY 70 SAGLVAAASPYVAMSVTAQAQELTAQVRAAAAYETAYGLTVPPVPPVIAENRAELMIL 129
Db 61 ASAAVAEAVAPYVAMSAQAQAQEAQATQARAANAFAFAATVPPLTAANRASLMQL 120
QY 130 IATNLGQNTPAIVNEAEYGENWAQDAAMFYAAATATATATATLTPFEAPDWTAGGL 189
Db 121 ISTNVFGQNTSAIAAAEAQYGENWAQDSAAAYAGSSASASA-VTFPSTPPQIANETAQ 179
QY 190 LEQAAYEEASDTAAA-NOLMNNVPOALQOLAQPTQGTTPSSKLGGLWK----- 237
Db 180 GTQAAAVATAAGTAQSTLTEMITGLPNALQSLTSPILQSS-NGPLSWLWQLLFGTNPFT 238
QY 238 -----TVSPHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVOTAA 290
Db 239 SISALLTDLQFYASFFNYTEGLPYFSGTGMNFIQAATL-GLIGSAAPAAVA-----AA 292
QY 291 QNGVRAVSSISLSSGSGGGVAAAGVAAAGVSLVSPQAAWAA-ANQAVTPAARALPLT 349
Db 293 GDAAGLPLGGMLG---GGPVAGLGNAAVSGKLSVPPVWSGGLPGSVTPGAAPLPVS 348
QY 350 SLTSAABERGQMLGGLPVQMGARAGGGLGVLRVPPRPVPMHPSPAAG 399
Db 349 TVSAAPAEAPGSLGGLPL----AGAGGAGAGP-RYGFRPTVMARPPFAG 393

RESULT 10
OS3939 PRELIMINARY; PRT; 411 AA.
AC OS3939;
DT 01-JUN-1998 (TReMBLrel. 06, Created)
DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE PPE-family protein.
GN RV1789 OR MT1838 OR MTV049.11.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;

```

RA Fleischmann R.D., Alland D., Eissen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwin M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.,
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL020201; CAAL7711.1; ALT_INIT.
 DR EMBL; AB007043; AAK46108.1; --
 DR PIR; G70929; G70929.
 DR TIGR; MT1838; --
 DR TubercuList; Rv1789; --
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Complete proteome.
 SQ SEQUENCE 411 AA; 40557 MW; 03C90B5E0590B7DA CRC64;

Query Match 25.9%; Score 779.5; DB 16; Length 411;
 Best Local Similarity 43.7%; Pred. No. 1.5e-31;
 Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;

QY 10 VDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVMGLTVGSWTGS 69
 DB 19 MDFGALPPEVNSVRMYAGPGSAPMVAAASAWNGLAELSSAATGYETVITQLSSEGWLGP 78

QY 70 SAGLMVAASPPYVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPPVIAENRAELMIL 129
 DB 79 ASAMAEAVAPYVAMSAQAQAEQATQARAAAAFAFAFAFAFAFAFAFAFAFAFAFAFA 138

QY 130 IATNLGONTPAIVANEABYGEWMAODAAAFYGAATATATATATATATATATATATATATAT 189
 DB 139 ISTNVFGONTSAIAAEAYGEWMAODSAAMYAYAGSSASASA-VTPFGTTPQIANPTAQ 197

QY 190 LEQAAVEERASDTAAA--NOLMNVFQALQQLAQPTGGTTPSKLGLWK----- 237
 DB 198 GTQAAAVATAAGTAQSTLTETMTGLPNAQSILTSPLQSS--NGPLSLWMLQILFGTNPFT 256

QY 238 -----TVSPHSPISNMVSMANNMSTNSGVSMNTLSMLKGFAPAAAAAQVATAA 290
 DB 257 SIGALLTDLQFYASFFNTBGLPYFYGIMGNFIQAKTL-GLIGSAAAPAAVA-----AA 310

QY 291 QNGVRAMSSLSGSLGSGGVAANLGRAASVGSLSVFOANAA--ANQAVTPAARALPIT 349
 DB 311 GDAAKGLPLGGLMG---GCPVAAGLGNAAASVGLSVPPVWSGSLPFGSVTPCAAPLPVS 366

QY 350 SLTSAERGGQMLGGLPVQCMGARAGGSLGVLRVPPRPPYMPHSPAAG 399
 DB 367 TVSAAPAPAPGSLGGLPL-----AGAGGAGAGP-RYGFRTVVARPPFAG 411

RESULT 11
 Q50320 PRELIMINARY; PRT; 361 AA.

AC Q50320; DB 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE 34kDa protein precursor.
 OS Mycobacterium paratuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1770;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JDB8/107;
 RA MEDLINE=96005449; PubMed=7921248;
 RA Cameron R.M., Stevenson K., Inglis N.F., Klausen J., Sharp J.M.;
 RT "Identification and characterisation of a putative serine protease
 RT expressed in vivo by Mycobacterium avium subsp paratuberculosis."
 RL Microbiology 140:1977-1982(1994).
 CC -!- SIMILARITY: CONTAINS 1 PDZ/DRH DOMAIN.
 DB EMBL; Z23092; CAA80638.1; --

DR PIR; S47170; S47170.
 DR GO; GO:0008233; P:peptidase activity; IEA.
 DR GO; GO:0004295; P:trypsin activity; IEA.
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 1.
 DR Pfam; PF00595; PDZ; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SMO0228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 KW Hydrolase; Protease; Serine protease; Signal.
 FT SIGNAL 1 38 POTENTIAL
 SQ SEQUENCE 361 AA; 35709 MW; 30FEF78FD6F3C411 CRC64;

Query Match 25.9%; Score 778.5; DB 2; Length 361;
 Best Local Similarity 66.0%; Pred. No. 1.4e-31;
 Matches 155; Conservative 28; Mismatches 37; Indels 15; Gaps 2;

QY 362 MGLGFLPQCMGARAGGSLGVLRVPPRPPYMPHSPAAGDIAPALSDODRFPALPLDP 421
 DB 15 LVGLTVVGLGLGSGVG-----LAPASA--APSGLALDRPADRLAPIDP 59

QY 422 SAMVAQGPQVMTINTKLYNNNAVAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQ 481
 DB 60 SAMVGQGPQVMTIDTKFYNNNAVAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQ 119

QY 482 TYGVVDVVGVDRTQVAVLQRLRAGGLPSAAGGVAVGEPVAMKSGGGGCTPRVPGR 541
 DB 120 TYAVDVVGYDRTQDIAVLQRLRAGGLPSAAGGVAVGEPVAMKSGGGGCTPRVPGR 179

QY 542 VVALGQTVQASDSLTGAETLNGLIQFDAAIQPGDSGPPVNGLGQVGNMTAAS 596
 DB 180 VVALNQSVSATIDLTGAQNLGLIQADAPIKFGDSGPPVNSAGQVIGVDVTAAT 234

RESULT 12
 Q99Q11 PRELIMINARY; PRT; 410 AA.

AC Q99Q11; DB 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Rv1808-like protein.
 GN MYC1808 OR OV1808.
 OS Mycobacterium microti.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1806;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MYC 94-2272, and OV254;
 RA Sivadon V., Heym B., Mazancourt P., Gaillard J.-L.L.;
 RT "PPE Rv1808 orthologue of Mycobacterium microti."
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF335180; AAK20894.1; --
 DR EMBL; AF335179; AAK20893.1; --
 DR InterPro; IPR000030; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 SQ SEQUENCE 410 AA; 40044 MW; 3B9DA3174655A5EA CRC64;

Query Match 25.2%; Score 756.5; DB 2; Length 410;
 Best Local Similarity 42.4%; Pred. No. 2.1e-30;
 Matches 182; Conservative 52; Mismatches 154; Indels 41; Gaps 11;

QY 9 MVDFGALPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVMGLTVGSWTGS 68
 DB 1 MLDGALPPEINSGRMYAGPGSGPLAAAWDAAALAAELYSAAASVGSSTIEGLTVAPMWG 60

QY 69 SSAGLMVAASPPYVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPPVIAENRAELMI 128

Db 61 PSITMAAAVYVAVISYTAQAAQAAKIAAGVYETAFAAATPPPPVIEANRALLMS 120
QY 129 LIATNLLGNTPAIAVNEAEYEMWAQDAAMFGVAAATATATATATLLPFEAPETMSAGG 188
Db 121 LVATNIFGNTPAIAATEAHYAEMWAQDAAMFGVAAATATATATATLLPFEAPETMSAGG 179
QY 189 LLEQAAVVEASDTAAA-----NQLMNNVFOALQQLAQPTQGTTPSKLGLWKTVSP 241
Db 180 TAAQSAVAAQAAGAAASSDITAQLSQLISLLPSTLIQSLA--TTATATSASAG--WDTV-- 233
QY 242 HRSPISNMYSMMNMTNSGV---SMTNTLSMLKGFAPAAAQAQVTAQNGVRAMS 298
Db 234 -LQSIITILIANTGPIYSITGLGAIPGGWLLTFGQIL-GLAQAQVGAALLGPKAAAGALS 291
QY 299 SLGSSLSGS-----SLGGGVAANLGRASVGLSVLPQAWAANQAQVTPAARALPITSLS 353
Db 292 PLAPLRGGYIADITPLGGGATGIAIYVGLSVLPQAWAANQAQVTPAARALPITSLS 351
QY 354 A-AERPGQWGLGGLPVQNGAGAGGL-----SGVLRYPPRYVPHSPAGDIAPFALS 407
Db 352 ALAAEPAGALFGEMALSSLAGRALAGTAVRSAGAAARV-----AGGSVTEVDAS 400
QY 408 QDRFADFFA 416
Db 401 TTTIIVIPA 409

RESULT 13

Q9CCV9 ID Q9CCV9 PRELIMINARY; PRT; 354 AA.
AC DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Probable secreted serine protease.
GN ML2659.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Dughey S., Feltwell T., Fraser A., Hamlin N.,
RA Holtroyd S., Hornsby T., Jagsels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus."
RL Nature 409:1007-1011(2001).
CC -/- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL; AL583926; CAC32191.1; --
DR PIR; A87242; A87242.
DR Leptoma; ML2659; --
DR GO; GO:0004295; F:trypsin activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001254; Peptidase S1.
DR InterPro; IPR008256; Peptidase_S1B_V8.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR PRINTS; PR00839; V8PROTEASE.
DR SMART; SM00228; PDZ; 1.

DR PROSITE; PS0106; PDZ; 1.
KW Hydrolase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 354 AA; 3265 MW; 612F23261BC9EA4A CRC64;
Query Match 25.1%; Score 754.5; DB 16; Length 354;
Best Local Similarity 73.3%; Pred. No. 2.2e-30;
Matches 151; Conservative 25; Mismatches 25; Indels 5; Gaps 2;
QY 391 VMHSPAGDIAPFALSQDRPADFPALPLDPSAMVAQVGVNNTKLYNNNAVAGATG 450
Db 27 VVPGS--ATPSGPTLALDRFSNRPLPLNPAAMVA---PQVYNISTRLYNSAVGAGTG 81
QY 451 IVIDPENGVLNNHVIAGATDINAFSVGSGTGVGVVVDRTQDVAVLQLRGAGGLPSA 510
Db 82 IVIDSSGVVLTNNHVISGATDISAFDVGNKGTGVGVVDRTQDVAVLQLRGASNLPTA 141
QY 511 AIGGVAVGVFVAMGNSGGGTTPRAVPGRVVALGTQVQASDLTGAEETLGLIQFDA 570
Db 142 VIGGDVAIGEPIVALGNTGGGGLPSVLPGRVVALNQTQVASEPLTGAQETLSLIQVDA 201
QY 571 AIQPGDSGGPVNGLGQVGMNTAAS 596
Db 202 PIKFGDSGGPVNVRGQVGMNTAAT 227
RESULT 14
O53957 ID O53957 PRELIMINARY; PRT; 409 AA.
AC DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE PPE-family protein.
GN RV1808 OR MT1855.1 OR MTV049.30.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H37Rv;
RC MEDLINE=96295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagsels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Elnolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains."
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL022021; CAA17729.1; --
DR EMBL; AE007044; AAK46129.1; ALT_INIT.
DR PIR; A70932; A70932.
DR TIGR; MT1856.1; --
DR TubercuList; RV1808; --
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Complete proteome.
SQ SEQUENCE 409 AA; 39917 MW; 1B15202BACF36379 CRC64;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:42:21 ; Search time 9.70351 Seconds
(without alignments)
3198.204 Million cell updates/sec

Title: US-09-597-796C-10

Perfect score: 3007

Sequence: 1 HHMHHHHMDVFGALPPPIN.....SGGPVNVGLGVGMNTRAS 596

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	ID	Description
1	1656.5	55.1	YD61_MYCTU	Q1031 mycobacteri
2	705	23.4	Y592_MYCTU	Q10813 mycobacteri
3	688.5	22.9	Y102_MYCTU	O53951 mycobacteri
4	448	14.9	YF48_MYCTU	Q10778 mycobacteri
5	445	14.8	Y442_MYCTU	P42611 mycobacteri
6	426.5	14.2	Y478_MYCTU	Q10540 mycobacteri
7	365	12.1	SRA_MYCLE	Q07297 mycobacteri
8	338	11.2	YU18_MYCTU	P31500 mycobacteri
9	334	11.1	Y096_MYCTU	Q10892 mycobacteri
10	333.5	11.1	YU21_MYCTU	Q10892 mycobacteri
11	262.5	8.7	YVTA_BACSU	O53268 mycobacteri
12	233.5	7.8	YX29_MYCTU	Q91911 bacillus su
13	225.5	7.5	YX28_ARATH	O06246 mycobacteri
14	225	7.5	DEGS_ECOLI	Q91u10 arabidopsis
15	225	7.5	DEGP_ECOLI	P31137 escherichia
16	224.5	7.5	DEGP_CHLEN	Q92670 chlamydia p
17	222	7.4	DEGP_HUMAN	O43464 homo sapien
18	222	7.4	DEGP_ARATH	O22660 arabidopsis
19	221.5	7.4	DEGP_MOUSE	P39099 escherichia
20	221	7.3	YH2A_MOUSE	O91iY5 mus musculu
21	221	7.3	DEGP_BRUSU	O34358 bacillus su
22	219.5	7.3	YXKA_BACSU	Q44597 brucella su
23	218.5	7.3	DEGP_BUCAI	P39668 bacillus su
24	217.5	7.2	YX25_MYCTU	P57322 buchnera ap
25	217	7.2	HTOA_HAEIN	O50703 mycobacteri
26	215	7.1	HTRA_LACHE	P45129 haemophilus
27	215	7.1	DEGP_EICGN	O824h7 lactobacill
28	213	7.1	DEGP_BRUNE	Q92jal rickettsia
29	211	7.0	DEGP_SALTY	O8Y532 brucella me
30	208.5	6.9	DEGP_ECOLI	P26982 salmonella
31	208.5	6.9	DEGP_CHLTR	P09376 escherichia
32	207.5	6.9	DEGP_BUCAP	P18584 chlamydia t
33	206.5	6.9	DEGP_CHLMU	O85291 buchnera ap
				Q9p197 chlamydia m

ALIGNMENTS

RESULT 1

ID	YD61_MYCTU	STANDARD;	PRT;	396 AA.
AC	Q11031;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	Hypothetical PPE-family protein Rv1361c/MT1406.			
GN	Rv1361C OR MT1406 OR MTCY02B10.25C.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;			
OC	Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.			
OX	NCBI_TaxID=1773;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=H37RV;			
RC	MEDLINE=98295987; PubMed=9634230;			
RA	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Horneby R., Jagels K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Oliver S., Osborne J., Krogg A., McLean J., Moule S., Murphy L.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence."			
RL	Nature 393:537-544(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CDC 1551 / Oshkosh;			
RC	MEDLINE=22206494; PubMed=12218036;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,			
RA	Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,			
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;			
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains."			
RL	J. Bacteriol. 184:5479-5490(2002).			
CC	-1- SIMILARITY: Belongs to the mycobacterial PPE family.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; Z75555; CAA99966.1; -			
DR	EMBL; AE007013; AAK45669.1; -			
DR	PIR; H70741; H70741.			
DR	TIGR; MT1406; -			
DR	Tuberculist; Rv1361c; -			
DR	InterPro; IPR000030; Microbac_PPE.			

Q52894 rhizobium m
Q05942 rickettsia
P54925 bartonella
Q93372 rattus norv
P83105 homo sapien
P15502 homo sapien
Q91a06 lactococcus
Q92743 homo sapien
P04985 bos taurus
Q91118 mus musculu
P83110 homo sapien
P44947 haemophilus

```

DR Pfam: PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 158 159 TA -> AT (IN REF. 2).
SQ SEQUENCE 396 AA; 40015 MW; 6AF60D7B5F668D0 CRC64;

Query Match 55.1%; Score 1656.5; DB 1; Length 396;
Best Local Similarity 85.1%; Pred. No. 5.5e-77;
Matches 338; Conservative 19; Mismatches 34; Indels 7; Gaps 3;

QY 9 MVDGALPPPEINARMYAGPGSASLVAAAKQWDSVASDLFSAASATQSVWGLTVGWSWG 58
DB 1 MVDGALPPPEINARMYAGPGSASLVAAAKQWDSVASDLFSAASATQSVWGLTVGWSWG 60
QY 69 SSAGLVAAASPVVWMSVTTAGQELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMI 128
DB 61 SSAGLVAAASPVVWMSVTTAGQELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMI 120
QY 129 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATALLPPEAPEMTSAGG 188
DB 121 LIATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATALLPPEAPEMTSAGG 180
QY 189 LLEQAAVEEASDTAAANQIMWNPQALQLOAQTOGTTTPSSKLGGLWKTVPSPHRSPIGN 248
DB 181 LLEQAAVEEASDTAAANQIMWNPQALQLOAQTOGTTTPSSKLGGLWKTVPSPHRSPIGN 240
QY 249 MVSMANNHMTNMGVSMNTNLTSSMLKGFAPAAAQAVOTAAQNGVRVSS-----LGSSL 304
DB 241 IVSMNNHVMNTNMGVSMNTNLTSSMLKGFAP-AAAQAVETAAQNGVQVAAQNGVSSLGSS 299
QY 305 GSSGLGGVAAANLGRAASVCSLSVPQAAANQAVTPAARALPLTSLTSAABEGPQMLG 364
DB 300 GSSGLGGVAAANLGRAASVCSLSVPQAAANQAVTPAARALPLTSLTSAABEGPQMLG 359
QY 365 GLPVGQMGARAG--GGLSGLVLRVPPPPVPMHPSPAAG 399
DB 360 GLPVGQMGARAG--GGLSGLVLRVPPPPVPMHPSPAAG 396

RESULT 2
YS92 MYCTU STANDARD; PRT; 408 AA.
AC Q10813;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein RV2892C/MT2959/MB2916c.
GN RV2892C OR MT2959 OR MTCY274.23C OR MB2916C.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773, 1765;
RN [1];
SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Teklaia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holtroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN [2];
SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=22206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,

```

```

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL J. Bacteriol. 184:5479-5490(2002).
RN [3];
SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duchoy S., Gordin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -I- SIMILARITY: Belongs to the mycobacterial PPE family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; 274024; CAA98377.1; -.
DR EMBL; AE007119; AAK47285.1; -.
DR EMBL; BX248344; CAD96603.1; -.
DR PIR; G70925; G70925.
DR TIGR; MT2959; -.
DR TubercuList; RV2892C; -.
DR InterPro; IPR000030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 56
FT POTENTIAL.
SQ SEQUENCE 408 AA; 41469 MW; 3E3D1F20D7827199 CRC64;

Query Match 23.4%; Score 705; DB 1; Length 408;
Best Local Similarity 41.8%; Pred. No. 5.2e-28;
Matches 182; Conservative 47; Mismatches 134; Indels 72; Gaps 14;

QY 10 VDFGALPPEINARMYAGPGSASLVAAAKQWDSVASDLFSAASATQSVWGLTVGWSWG 69
DB 1 MDFGVLPPPEINSGRMVYAGPGSGPWWMAAAAADSLAELGLAAGGYRLAISLGTGAYWAGP 60
QY 70 SAGLVAAASPVVWMSVTTAGQELTAQVRVAAAAYETAYGLTVPPPPVIAENRAELMI 129
DB 61 AAASWVAATVPYVAVLSTAGQAEQAGQAAQAAAYELAFAMTVPPPPVIAENRAELMI 120
QY 130 LATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATALLPPEAPEMTSAGG 189
DB 121 VATNFFGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATALLPPEAPEMTSAGG 179
QY 190 LEQAAA-----VEEASDTAAANQIMWNP-----VP--QALQO-LAQPTQGTTPSSKL--- 232
DB 180 AQAAAATVSTVTPPLATTAAPVQLQQLSLSLTPWYSLAQQLWLAENLLGLTPPNRTIV 239
QY 233 -----GGLWKTVPSPHRSPISNMVMNNHMTNMGVSMNTNLTSSMLKGFAPAAA 283
DB 240 RLGLISYFDEGL-----LQFEASLAQQAIFGTPGGAG--DSGSSVLDSWGPTIFA 287
QY 284 QAVQTAAQNGVRMSSL--GSSLSG-----SGLGGVAAANLGRAASVCSLS 327
DB 288 -----GPRASPSVAGGAGVGQVTPQPYWYALDRESIGSVSAALGKGSAGSLS 338
QY 328 VPQAAANQAVTPAARALP---LTSLSAERGGMGGLPVGQMGARAGGGLSGVLR 384
DB 339 VPPDWAARARWANPAMWLPDGDVTLRGTAENA---LLRGFFPNASAGQSTGGGF--VHK 393
QY 385 VPPPEYVPMHPSPAAG 399
DB 394 YGFRILAVMQRPFFAG 408

```

```
RESULT 3
ID Y102 MYCTU STANDARD; PRT; 463 AA.
AC OS3951;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rv1802/MT1851/Mb1830.
GN Rv1802 OR MT1851 OR Mb1830.24 OR Mb1830.
OS Mycobacterium tuberculosis, and
OS Mycobacterium bovis.
OC Corynebacteriaceae; Actinobacteridae; Actinomycetales;
OC Bacteria; Actinobacteria;
OX NCBI_TaxID=1773, 1765;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RA "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=2206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
RL laboratory strains.";
RJ J. Bacteriol. 184:5479-5490 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=M.bovis; STRAIN=AF2122/97;
RX MEDLINE=22709107; PubMed=12788972;
RA Garnier T., Eiglmeyer K., Camus J.-C., Medina N., Mansoor H.,
RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
RT "The complete genome sequence of Mycobacterium bovis.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882 (2003).
CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL022021; CAAL7723.1; -
DR EMBL; AE007044; AAK46123.1; -
DR EMBL; BX248340; CAD94533.1; -
DR FIC; C70931; C70931.
DR TIGR; MT1851; -
DR TuberculList; Rv1802; -
DR InterPro; IPR00030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 401 401 S -> L (IN REF. 2).
```

```
SQ SEQUENCE 463 AA; 46021 MW; EB64828BF09FA551 CRC64;
Query Match 22.9%; Score 688.5; DB 1; Length 463;
Best Local Similarity 42.8%; Pred. No. 48-28;
Matches 174; Conservative 48; Mismatches 136; Indels 49; Gaps 11;

QY 10 VDFGALPPEINSARMYAGPGSALVAAAQMWDSVASDLFSAAAFQSVVWGLTVGSWIGS 69
DQ 1 MDFGVLPPEINSGRMYAGPGSGPMLAAAAMADGLATELQSTADYGSVISVLT-GVWSGQ 59
QY 70 SAGLMVAASPYVAMSVTAQAELETAQVTVAAATYGLTVPPPIVIAENRAELMIL 129
DQ 60 SSGTMAAAAPYVAMSVTAQAELETAQVTVAAATYGLTVPPPIVIAENRAELMIL 119
QY 130 IATNLLGONTTALVNEAEYGEWMAQDAAMAFVAAAATATATATLLPPEEPAPMTSAGCL 189
DQ 120 AATNIFGQNTGATIAAAEAARYAEMWAQDAAMAFVAAAATATATATLLPPEEPAPMTSAGCL 178
QY 190 LEQAAVEEASDTAAANQNMNVFQALQLOAQPTGGTTPPSKLGGLMKTVS--PHRSP- 246
DQ 179 ATQGVAVAAQVAGAGAGN-ARSLVSEVLEFLA--TAGTNYNKTVASLMNAVTVGVYASSV 235
QY 247 -----SNMYSMANNHSMNTNSGVSMNTLSSMLKGFAPAAAQAQVOTAAQNGVRA 296
DQ 236 NSMLGLGFAESKRVLPANDTVISIFGMVQFKFNPVPFNPDLIPK----- 283
QY 297 MSSLGSSLG-----SSGLGG---GVANLGRAASVGSLSVPQAAANAAQVTPAARALPL 348
DQ 284 -SALGAGLGLRSALSSGLGSTAPALSAGASQAGSVGMSVPSVPSAAATPAIRTVAAVFS 342
QY 349 TSLTS--AERGPQGL-----GGLPVQMGARAGGGLSGSLVRV 385
DQ 343 TGLQAVPAAAISEGLSLQMALASVAGGALGAAARATGTFGLGGRV 389

RESULT 4
YF48_MYCTU
ID YF48 MYCTU STANDARD; PRT; 678 AA.
AC Q10778;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical PPE-family protein Rv1548c/MT1599.
GN Rv1548C OR MT1599 OR MYC48.17.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeyer K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RX MEDLINE=2206494; PubMed=12218036;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
```

laboratory strains.",
 RT J. Bacteriol. 184:5479-5490(2002).
 RL -!- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z74020; CRA98335.1; -;
 DR EMBL; AE007026; AAK45866.1; ALT_INIT.
 DR PIR; A70762; A70762.
 DR TIGR; MT1599; -;
 DR TubercuList; Rv1548c; -;
 DR InterPro; IPR000030; Microbac_PPE.
 DR InterPro; IPR002989; Mycobac_Pentapep.
 DR Pfam; PF01469; Pentapeptide_2; 1.
 DR Pfam; PF0823; PPE; 1.
 KW Hypothetical protein; Transmembrane; Complete proteome.
 FT TRANSMEM 14 34 POTENTIAL.
 FT TRANSMEM 180 200 POTENTIAL.
 FT CONFLICT 258 258 D -> G (IN REF. 2).
 SQ SEQUENCE 678 AA; 66736 MW; 209F1593D52533A2 CRC64;
 Query Match 14.9%; Score 448; DB 1; Length 678;
 Best Local Similarity 26.5%; Pred. No. 8.1e-16;
 Matches 155; Conservative 70; Mismatches 209; Indels 150; Gaps 20;
 QY 10 VDFGALPEINSMRYAGPGSASLVAAQWDSVASDLFSAASAFQSVWGLTVGSGIGS 69
 DB 1 MNFSLVPEINSMALMPAGPGPMLAASAWTGLAGDLGSAASFSAVTSQLATGSGWGP 60
 QY 70 SAGLMAAASPYVAMSVTAQAEILTAQVRAAAAYETAYGLTVPPVIAENRAEIMLIL 129
 DB 61 ASAAATGVAASVYARWLTTAAQAQAAQAAQAAQAAQAAQAAQAAQAAQAAQAAQAA 120
 QY 130 IATNLGONTAIAVNEAEYEMQAQAAAFMFGAAATATATATATATATATATATATATAT 189
 DB 121 VASNLGONAIAVNEAEYEMQAQAAAFMFGAAATATATATATATATATATATATATAT 172
 QY 190 LEQAAAEASDSTAANQLMN--NVPOALQOLAQFTQGTTPSSKLGGLWKTVPSPHSPI 247
 DB 173 --SAAATPGGAVIIAGPFLDLGNV-----TIGGF----- 200
 QY 248 MNSVMANNHMTNSGYSMNTLSSMLKGFAPAAQAAQAAQAAQAAQAAQAAQAAQAAQAA 307
 DB 201 ---NLASNLGLNLG-----SFPNGSANTGVSNLGNANTGDLNLGSGNIGTGY 245
 QY 308 GLGGGVAANLG-RAASVGSLSVPQAAANQAVTPAARALPLSLTSAABRGQOMLG-- 364
 DB 246 NLGGNTGDLNPDSGNTGLN-----WQSGN-----IGSYNLGGN-LGSY 285
 QY 365 GLFVGQMG-ARAGGSLGVLEVPVPPVYMPHSPHSPAGDIAPALPSQDRFAPALPLDPSA 423
 DB 286 NLGSGNTGDTNFGGNTGNLNVG----- 308
 QY 424 MVAQVGQPVNNTKLGYNNAVAGAGTGIVIDPN---GWLTNNHVIAGATDINAFSVGSG 480
 DB 309 -----GGNTGNSNFGNGTGNVFNNGTGTNFGSGNLGSGN---LGFNGKSHNIGFG 360
 QY 481 QTYGVVVVYDRTQDVAVLQIRGAGGLFSA--IGGVAVGEPVPMVNGS-----GGGGT 534
 DB 361 NS-GNNNIGFGLTGDNIQI-----GFCALNSGSGNLGFGNS-GNGNIGFNSGNNIGMGN 414
 QY 535 PRAYPGRVVALGQTVQASDSLTGAETLNGLIQFDAAIQPCDSG 578
 DB 415 GNGVGLSVFEGSSAERSGSGFNGSGELSTGI-----GNSG 449

RESULT 5

Y442_MYCTU
 ID Y442 MYCTU STANDARD; PRT; 487 AA.
 AC F42611; OS3727;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PPE-family protein Rv0442c/MT0458.
 GN Rv0442c OR MT0458 OR MT037.06C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]_SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RC STRAIN=Ersmann;
 RX MEDLINE=87137260; PubMed=3029018;
 RA Shinnick T.M.;
 RT "The 65-kilodalton antigen of Mycobacterium tuberculosis.";
 RL J. Bacteriol. 169:1080-1088(1987).
 RN [2]_SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sultston J.E., Taylor K., Whitthead S., Barrell B.G.;
 RA "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [3]_SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischnann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Ustebach T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RL laboratory strains."
 CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; ML5467; AAA88235.1; ALT_INIT.
 DR EMBL; AL021932; AAK17399.1; -;
 DR EMBL; AE006948; AAK44681.1; -;
 DR PIR; C70830; C70830.
 DR TIGR; MT0458; -;
 DR TubercuList; Rv0442c; -;
 DR InterPro; IPR000030; Microbac_PPE.
 DR InterPro; IPR002989; Mycobac_Pentapep.
 DR Pfam; PF01469; Pentapeptide_2; 5.
 DR Pfam; PF0823; PPE; 1.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 40 40 E -> K (IN REF. 2).
 FT CONFLICT 96 96 I -> T (IN REF. 1).
 FT CONFLICT 211 211 G -> GNNIG (IN REF. 1).
 SQ SEQUENCE 487 AA; 47247 MW; 97234D5B316C8C7F CRC64;
 Query Match 14.8%; Score 445; DB 1; Length 487;


```

[3]
SEQUENCE OF 160-374 FROM N.A.
RN RP RC
SEQUENCE=Isolate 50410;
RA Patki A.H., Dale J.W.;
Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the mycobacterial ppe family.
CC -1- CAUTION: In strain Oshkeh the gene for this protein is
interrupted in position 307 by an IS6110 element.
CC -1- CAUTION: Was originally (Ref.3) thought to be a dihydrofolate
reductase.
CC -1- CAUTION: Ref.3 sequence differs from that shown due to frameshifts
in positions 294; 337 and 355.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL021287; CAA16103.1; -.
DR DR EMBL; AB007129; AAK47427.1; ALT_SEQ.
DR EMBL; AB007129; AAK47430.1; ALT_SEQ.
DR EMBL; X59271; CAA41961.1; ALT_FRAME.
DR FIR; E70857; E70857.
DR TIGR; MT3098; -.
DR TIGR; MT3101; -.
DR TubercuList; RV3018c; -.
DR InterPro; IPR00030; Microbac_PPE.
DR Pfam; PF00823; PPE; 1.
DR HypoThetical protein; Complete proteome.
KW SEQUENCE 434 AA; 43029 MW; 41D673C4BD389DD6 CRC64;
SQ
Query Match 11.2%; Score 338; DB 1; Length 434;
Best Local Similarity 27.2%; Pred. No. 1.8e-11;
Matches 123; Conservative 54; Mismatches 211; Indels 64; Gaps 12;

QY 14 ALPPEINSRMYAGPGSASIVAAQWDSVASDLPFAASAQFQVVMGLTVGSMIGSSAGL 73
DB 8 ASPPEVHSALLSAGPGFGLQAAAGWSALSASYAAVQELSVVVAAGVWGQGSABL 67
QY 74 MVAASPYVAMSVTGAQBELTAAQVVRVAAAAYETAYGLTVPPPVIAENRAELMILIAIN 133
DB 68 FVAAYYPVYVAVLQVQASDASAAAGHEAAAAGVVCALAEMPTLPFLAANHLTHAVLVATN 127
QY 134 LIGQNPATAVNBAEYGEWMAQDAAMFGVAARATATATLTFPEAPEMTSAGGLEQA 193
DB 128 FGINPTIPALNEADYRVMVQQAATVMSAYEVUGAALVATHTGPAPVVRKP----- 181
QY 194 AAVEASDTAAAN-----QLMNNVPQALQLAQTQGTTPSSKGLGLWKTVPSPH 242
DB 182 --ANEAASNAVAATITPPFPWHEIVQFLEETFAAYDQVLSALLSELPA--VAVWVFLQFVD 237
QY 243 ---RSPISNVSMNNHMSNTSGVMTNTLSSMLKGFAPAAAAQAVOTAAQNGVRAMS 299
DB 238 ILGFNIIGFIITLASHNAQLITEPAINASVAVGLLYAIA-GVTDIVVEWVIGNLFQVVP 296
QY 300 LGSSL-----GSSGLGG--GVAA--NLGRAASVGSLSVPQAWAANAQVTPAAPA 345
DB 297 LGGFLLGALAAA VVPGVAGIAGVAGLAALPAVGAAAGAPAAALVGSVAPVSGGVSPQAR- 355
QY 346 LPLTSLTSAERGGQMLGGLPVQWGAGAGG--GLSGVLRVPPRPVYVPHSPAAQDIAP 403
DB 356 -----LVSAVEPAPASTSVSVLASDRGAGALGFPVGTAG-----KESVQOPAG 397
QY 404 PALSQDRFADFPAALPDPMSVMAVQGVQVYNI 435
DB 398 LTVLADFEFGDGPVPMPLGWSW----GPDVLGV 425

```

RESULT 9
Y096 MYCTU

ID	Y096_MYCTU	STANDARD;	PRT;	463 AA.
DI	Q10892;			
DT	AC			
DT	01-OCT-1996	(Rel. 34, Created)		
DT	01-OCT-1996	(Rel. 34, Last sequence update)		
DT	10-OCT-2003	(Rel. 42, Last annotation update)		
DE	Hypochemical PPE-family protein Rv0096/MtO105.			
GN	Rv0096 OR MtO105 OR MTCy251.i15.			
OS	Mycobacterium tuberculosis.			
OC	Bacteria; Actinobacteria; Actinomycetales;			
OC	Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.			
NCBI_TaxID=1773;	[1]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=R37R;			
RP	MEDLINE=G98295987; PubMed=9634230;			
RX	Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,			
RA	Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,			
RA	Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,			
RA	Davies K., Devlin K., Fellwell T., Gentles S., Hamlin N., Holroyd S.,			
RA	Hornaby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,			
RA	Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,			
RA	Rutter S., Seeger K., Skelton S., Squares S., Squares R.,			
RA	Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;			
RT	"Deciphering the biology of Mycobacterium tuberculosis from the			
RT	complete genome sequence.";			
RT	Nature 393:537-544(1998).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RP	STRAIN=CDC 1551 / Oshkosh;			
RX	MEDLINE=22206494; PubMed=12218036;			
RA	Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,			
RA	Peterson J.P., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,			
RA	Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,			
RA	Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,			
RA	Bisai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;			
RT	"Whole-genome comparison of Mycobacterium tuberculosis clinical and			
RT	laboratory strains.";			
RL	J. Bacteriol. 184:5479-5490(2002).			
CC	-!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).			
CC	-!- SIMILARITY: Belongs to the mycobacterial PPE family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to licenseseib-sib.ch).			
CC				
DR	EMBL; Z74410; CAAG98932.1; -.			
DR	EMBL; AB006922; AAK44327.1; -.			
DR	PIR; H70750; H70750.			
DR	TIGR; MT0105; -.			
DR	Tuberculist; Rv0096; -.			
DR	InterPro; IPR000030; Microbac_PPE.			
DR	Fram; PR00823; PPE; 1.			
KW	Hypothetical protein; Transmembrane; Complete proteome.			
FT	TRANSMEM 3 23 POTENTIAL.			
FT	TRANSMEM 88 108 POTENTIAL.			
FT	TRANSMEM 112 132 POTENTIAL.			
FT	TRANSMEM 216 236 POTENTIAL.			
FT	TRANSMEM 245 285 POTENTIAL.			
FT	TRANSMEM 276 296 POTENTIAL.			
FT	TRANSMEM 323 343 POTENTIAL.			
FT	TRANSMEM 419 439 POTENTIAL.			
SEQ	SEQUENCE 463 AA; 46894 MW; 42D9D66A033D0DD8 CRC64;			
	Query Match			
	Best Local Similarity 11.1%; Score 334; DB 1; Length 463;			
	Matches 138; Conservative 63; Mismatches 197; Indels 146; Gaps 17;			
QY	14 ALPPPSINARMYAGEGSAALVAAAQWMDSLFSAASAFQSVVGLTVGSWGSSAGL 73			

Qy 14 ALPPEINSARMYACPGSALSILVAAQWDSVASDLFSAASAFQSVWGLTVGSWIGSSAGL 73

Db 2 AIPPEVHSGLLSAGCGPGSLLVAAQQWQELSDQYALACAEGLGQLGEVQVSSNQGTAAQ 61
 QY 74 MVAAGPYVAVMVSVTAGQAEITAAQVRAAAAYETAYGLTVPVPIAENRAELMILIAIN 133
 Db 62 YVAAHGFYLAWEQTAINSVTAQVRAAAAYCSALAAAMPTPAELAAHAIHGVLIATN 121
 QY 134 LLGQNTFAIAVNEAEYGEVMAQDAAMFGVAAATATATATATLTPFEAPMTSAGLLGQA 193
 Db 122 FFGINTVPIALNEADYVRVWLQAADTAAAYQVADAATVAVPSTQAPPPIRAPGG 176
 QY 194 AAVEASDT-----AAANQNMNVPOALQQAQTO-----CTTPSSKL----- 232
 Db 177 -----DAADTRDLVLSIGQLIRI-----LDFIANPYKVFLEFFEQFGFSAVTVLALVAL 229
 QY 233 ---GGLW-----KTVSPHRSPISNMVSMANNHMTNSGVMNTNLSMLK-----GF 277
 Db 230 QLYDFLWYPYVYVSGLLLPFFTP-----TSLALTALSALIHLLNPPAGL 275
 QY 278 AAAAAQVTAQVRAQNGVRAMSSLSGSLGSGGGVAAVNGRAASVGSLSVQAAANQ 337
 Db 276 LPTAAA-----IGPDQMGANLA-----V 294
 QY 338 AVTPAARALPLTSLTSAAGPGQMLGGLFVQMGARAGGGLGVLRVPPRYVMPHSPA 397
 Db 295 AVTPATAAVP-----GGSPPTSNPAPAPSSNSVGSASAAFGI---SVA 335
 QY 398 AGDIAPPALSQDRPA--DPPALELDPMSVAVQVGPQVNVNTKLGYNNAVAGTGVIDP 455
 Db 336 VPGLEPPGVSGPKAGTKSPDPTAATLATAAGARPGLARHRRKRSQVGV-----IRGY 390
 QY 456 NGVVLTNHVIAGATDINAPSVSGSGTGYVDVGVYDRTQDVAQLRGAGGLPSAAIGGG 515
 Db 391 RDEFLDATAVDAATVPAPANNAG--SQAGTILGF-----AGTAPTIS---G 433
 QY 516 VAVG 519
 Db 434 AAAAG 437

RESULT 10
 YU21_MYCTU STANDARD; PRT; 435 AA.
 AC O53268; O53269;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical PEB-family protein RV3022c/RV3022c/MT3106.
 GN RV3022c/RV3022c OR MT3106 OR MT3106.35C/MT3106.36C.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=9825987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby R., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrall B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 complete genome sequence";
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,

RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Deicher A., Ufferback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains";
 RL J. Bacteriol. 184:5479-5490 (2002).
 CC -!- SIMILARITY: Belongs to the mycobacterial PPE family.
 CC -!- CAUTION: Ref.1 sequence differs from that shown due to a
 frameshift in position 82.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 DR EMBL; AL021287; CAAL16106.1; ALT_FRAME.
 DR EMBL; AL021287; CAAL16107.1; ALT_FRAME.
 DR EMBL; AE007129; AAK47435.1; -.
 DR TIGR; MT3106; -.
 DR Tuberculan; RV3021c; -.
 DR Tuberculan; RV3022c; -.
 DR InterPro; IPR000303; Microbac_PPE.
 DR Pfam; PF00823; PPE; 1.
 KW Hypothetical protein; Complete proteome.
 FT CONFLICT 299 299 G -> A (IN REF. 2).
 FT CONFLICT 317 320 LAGV -> VTGL (IN REF. 2).
 FT CONFLICT 326 326 L -> V (IN REF. 2).
 SQ SEQUENCE 435 AA; 42876 MW; 3B157643EAA8484A CRC64;
 Query Match 11.1%; Score 333.5; DB 1; Length 435;
 Best Local Similarity 25.3%; Pred. No. 3e-10;
 Matches 123; Conservative 59; Mismatches 172; Indels 133; Gaps 14;
 QY 14 ALPPEINARMYAGPGSASLVAAQWMDVSDLSAASAFQSVWGLTVGWSIGSAGL 73
 Db 8 ASPPEVHSHALLSAGPGPSLQAAAGWSALSAAVAAQELSVVAAVAGVWQGPSAEL 67
 QY 74 MVAAGPYVAVMVSVTAGQAEITAAQVRAAAAYETAYGLTVPVPIAENRAELMILIAIN 133
 Db 68 FVAAVPEVAVLWVQASADSAAGAEHAAAGVVCALAEMPTLPFLAANHLTHAVLATN 127
 QY 134 LLGQNTFAIAVNEAEYGEVMAQDAAMFGYAA----- 165
 Db 128 FFGINTVPIALNEADYVRVWVQAATVMSAYEAVVGAALVATPHTGPAVIVKFGANEASN 187
 QY 166 ATATATATLTPFE-----APEMTSAGLLGQAAVEASDSTAANQNMNVPOALQ 217
 Db 188 AVAAATITPPFFGLAKFLEMAQAQFTEVGELINKSAEAWAVGFVELITGLVNEP----- 243
 QY 218 QLAQPTQGTTPSSKLGGLWKTVPSPHRSPISNMVSM-----ANNH 256
 Db 244 -----WLW-----LTCMIDMFPATVGFALGVFLVPLLEFAVLE 278
 QY 257 MSMTNSGVMNTNLS-----MLKGFAPAAAQAQVQTAONGVRAMSSLSGSLGSSGL 309
 Db 279 LAILSIGMIISNIFGAIFVLGALLGALAAVVGVLGAGVAGLAAALPAVGAAGAP-- 336
 QY 310 GGGVAAVNLGRAASV-GSLSVFQAAVAAVQAATPAARALPLTSLTSAABERGPGMGLGLPV 368
 Db 337 -----AALVGSVAVPGVGVVQFQALVS--AVEPAPASTSVSL--ASDRGAGAL--GF-V 385
 QY 369 GQMGARAGGGLGVLRVPPRYVMPHSPAAAGDIAPPALSQDRPADFPALPLDPSAMVAVQ 428
 Db 386 GTAGKESVQAPAGL-----TVLADEFQDGPVPMPLPGSM----- 419
 QY 429 GPQVNI 435
 Db 420 GPDLVG 426

```

RESULT 11
ID YVTA_BACSU STANDARD; PRT; 458 AA.
AC Q9R9I1; O35021; O35039;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Probable serine protease yvta (EC 3.4.21.-).
GN YVTA OR YVTB OR BSU33000.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=20158875; PubMed=10692364;
RA Noone D., Howell A., Devine K.M.;
RT "Expression of ykda, encoding a Bacillus subtilis homologue of HtrA,
is heat shock inducible and negatively autoregulated.";
RL J. Bacteriol. 182:1592-1599 (2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98015415; PubMed=933931;
RA Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;
RT "Sequencing of regions downstream of addA (98 degrees) and citG (289
degrees) in Bacillus subtilis.";
RL Microbiology 143:3305-3308 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moser I., Albertini A.M., Alloni G.,
Azevedo V., Bertero M.G., Bessieres P., Solotkin A., Borcherdt S.,
Boriss R., Boursier L., Brans A., Braun M., Brignell M., Bron S.,
Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,
Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadaie Y.,
Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
Sorokin A., Tacconi E., Takagi T., Takahashi H., Takenaru K.,
Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis.";
RL Nature 390:249-256 (1997).
RN [4]
RP TRANSCRIPTIONAL REGULATION.
RC STRAIN=168;
RX MEDLINE=20576168; PubMed=11133960;
RA Noone D., Howell A., Collier R., Devine K.M.;
RT "ykda and yvta, HtrA-like serine proteases in Bacillus subtilis,
engage in negative autoregulation and reciprocal cross-regulation of
ykda and yvta gene expression.";
RL J. Bacteriol. 183:654-663 (2001).
CC -!- FUNCTION: May be involved in processing, maturation, or secretion
of extracellular enzymes.

```

```

CC -!- SUBCELLULAR LOCATION: Membrane-bound (Potential).
CC -!- INDUCTION: Induced by heat shock during exponential growth and by
CC heterologous amylases at the transition phase of the growth cycle.
CC Negatively regulates its own expression.
CC -!- MISCELLANEOUS: Inactivation results in compensating overexpression
CC of htrA, especially during stress conditions.
CC -!- SIMILARITY: Belongs to peptidase family S2C.
CC -!- SIMILARITY: Contains 1 P02/DHR domain.
CC -!- CAUTION: Ref.2 and Ref.3 sequences differ from that shown due to
CC frameshifts in positions 87 and 246 that produce two separate
CC ORFs.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF188296; AAF03153.1; -
CC EMBL; Z93941; CAB07968.1; ALT_FRAME.
CC EMBL; Z93941; CAB07969.1; ALT_FRAME.
CC EMBL; Z99120; CAB15290.1; ALT_FRAME.
CC Subtilist; BG14155; yvta.
CC InterPro; IPR009003; Cys_Ser_trypsin.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR001254; Peptidase_S1.
CC InterPro; IPR001940; Peptidase_SIC.
CC Pfam; PF00595; PDZ; 1.
CC Pfam; PF00083; trypsin; 1.
CC PRINTS; PR00834; PROTEASES2C.
CC SMART; SM00228; PDZ; 1.
CC PROSITE; PS0106; PDZ; 1.
CC Hydrolase; Protease; Serine protease; Heat shock; Transmembrane;
CC Complete proteome.
CC DOMAIN 1 71 CYTOPLASMIC (POTENTIAL).
CC TRANSMEM 72 92 POTENTIAL.
CC DOMAIN 93 458 EXTRACELLULAR (POTENTIAL).
CC DOMAIN 356 440 PDZ.
CC ACT_SITE 187 187 CHARGE RELAY SYSTEM (POTENTIAL).
CC ACT_SITE 217 217 CHARGE RELAY SYSTEM (POTENTIAL).
CC ACT_SITE 298 298 CHARGE RELAY SYSTEM (POTENTIAL).
CC SEQUENCE 458 AA; 48717 MW; 77551045A865ASCD CRC64;
CC -----
CC Query Match 8.7%; Score 262.5; DB 1; Length 458;
CC Best Local Similarity 29.1%; Pred. No. 1.2e-06;
CC Matches 85; Conservative 52; Mismatches 116; Indels 39; Gaps 14;
CC -----
CC QY 326 LSVPEQAAANQAVTPAARALPLTSLTSAERGGQMLGGI-PVGQMGARAGGGLS-GVL 383
CC Db 38 LDAPVTEAGRQ---ETASALEWEKQETAVKKEKRRRAWLSPI--LGGIIGGLMLGI- 91
CC QY 384 RVPRPYVPHSPAGDIAPALPSQDRFADFPALPLDPSA---MVAQVGQVYVNI-NTK 438
CC Db 92 ----APVLPDSDQATETA-SANKQVQSDNFTTAPITNASNIADWVDELTPTVIGISNIQ 146
CC QY 439 LGYNAV-----GAGTGIVI--DPNGVLTNNHVIAGATDINAFSVGSQTY 483
CC Db 147 TSQNNFTGTGGSSSESESGTSGVIFKQSDKXAYITNNHVVEGANKLTV-TLYNGETE 205
CC QY 484 GVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGG--VAVGEPVWAMGNGSGGGTTPRAVPR 541
CC Db 206 TAKLVGSDTTITDLAVLEISGNVKNVYASFGDSQSLRTGKVIAGINPLGQFSQSTVTOGI 265
CC QY 542 VVALGQTVQASDSLGTGAETTINGLIQFDAAIQPDGSGGPPVNGLGQVGVNWT 593
CC Db 266 ISGLNRTIDV-DTTQGTVEV--NVLTQDAAINFGSGGLPINASGOVIGINS 314
CC -----
CC RESULT 12
CC YI29 MYCTU
CC ID YI29_MYCTU STANDARD; PRT; 178 AA.

```


Db 41 VSVKVTQDKNHLHVLAVKSPVSTTRILLTSLFNLFCNPSRYLSALALGD----- 93
QY 378 GLSGVLRVPRPVYVMPHSPHSPAGDIAPPALSDQRFADFPALPDPS-AMVAQVGPQ-----V 432
Db 94 -----PSVATVEDVSTV-----FPAGPLFPTTEGRIVQLFKENTYSV 130
QY 433 VNI-----NTKLGYNNAV-----GAGTGIVDPNGVVLTHNVVIAGATDIN----- 473
Db 131 VNIFFVTLRPLQKMTGVBIPEGNSGVVMDQGGYIVTNYHYVHGNALSRNPSFGDWGRV 190
QY 474 --AFSGVSGQTVGVDDVGVYDRTQDVAVLQLRGAGGLPSAAI-----GGGVAVGEVVA 524
Db 191 NILASGVQKNEKLVGADRAKDLAVKV-----DAPETLLKPKVQGSNSLKVGQOCLA 246
QY 525 MNSGSGGQGTTPRAVGRVVALQVQASDLSLGAETLNGLLQFPAATQPDGSGPVPVNG 584
Db 247 IGNPFGFDHT--LTGVGVISGLNRDI---FSQTGV--TIGGGTQDAAINPGNSGGLDLS 299
QY 585 LGQVGVGMNTA 594
Db 300 KGNLIGINTA 309

RESULT 14
DEGS ECOLI
ID DEGS ECOLI STANDARD; PRT; 355 AA.
AC P31137; 1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protease degs precursor (EC 3.4.21.-).
GN DEGS OR HHOH OR HHRH OR B3235 OR Z4594 OR ECS4108.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RA Bass S., Gu Q., Gôddard A.;
RL Submitted (OCT-1994) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=96165272; PubMed=8576051;
RA Waller P.R., Sauer R.T.;
RT "Characterization of degQ and degS, Escherichia coli genes encoding
RL J. Bacteriol. 178:1146-1153(1996).
[3]
RN SEQUENCE FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[4]
RN SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RL Nature 409:529-533(2001).
[5]

RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kunara S., Siba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
[6]
RN PRELIMINARY SEQUENCE OF 82-355 FROM N.A.
RX MEDLINE=88105815; PubMed=3322223;
RA Vogel R.F., Entian K.-D., Mecke D.;
RT "Cloning and sequence of the mdh structural gene of Escherichia coli
coding for malate dehydrogenase.";
RL Arch. Microbiol. 149:36-42(1987).
[7]
RN IDENTIFICATION
RP Bazan J.F., Fletterick R.J.;
RA "Structural and catalytic models of trypsin-like viral proteases.";
RT Semin. Virol. 1:311-322(1990).
CC -I- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -I- SIMILARITY: Belongs to peptidase family S2C.
CC -I- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U15661; AAC43993.1; -;
DR EMBL; U32495; AAC44006.1; -;
DR EMBL; U18997; AAA58037.1; -;
DR EMBL; AE000402; AAC76267.1; -;
DR EMBL; AE005551; AAG58363.1; -;
DR EMBL; AP002564; BAB37531.1; -;
DR EMBL; M24777; -; NOT_ANNOTATED_CDS.
DR PIR; D91142; D91142.
DR MEROPS; S01.275; -;
DR EcoGene; EGI1652; hoh.
DR InterPro; IPR009003; Cys_Ser_trypsin.
DR InterPro; IPR001478; PDZ_Ser_trypsin.
DR InterPro; IPR001254; Peptidase_S1.
DR InterPro; IPR001940; Peptidase_S1C.
DR Pfam; PF00595; PDZ; 1.
DR PRINTS; PR00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR PROSITE; PS00228; PDZ; 1.
DR PROSITE; PS0105; PDZ; 1.
KW Hydrolyase; Serine protease; Periplasmic; Signal; Complete proteome.
FT SIGNAL 1 28
FT CHAIN 29 355
FT DOMAIN 281 326
FT ACT_SITE 96 96
FT ACT_SITE 126 126
FT ACT_SITE 201 201
FT ACT_SITE 253 253
FT CONFLICT 307 307
SQ SEQUENCE 355 AA; 37581 MW; D091B4D65E8FE1CC CRC64;
Query Match 7.5%; Score 225; DB 1; Length 355;
Best Local Similarity 30.9%; Pred.No. 7e-05;
Matches 67; Conservative 35; Mismatches 73; Indels 42; Gaps 9;
QY 396 PAAGDIAPPALSDQRFADFPALPDPSAMVAQVGPVNTKLGYNNAVGA----- 447
Db 24 PSLRSINP--LSTPQFDSITDTPASYNLAVRRAAPAVNV-----YNRGLNTNSHQLI 76

QY 448 --CTGIVIDPVGWLTNNHVIAGATDINAFSVSGQTYGVVDVVGDRDQDVAVLQGA 504
 DB 77 RTLGSGVIMQORGVIINKVINDADQI-IVAQDGRVFALLVGSLSLTLAVLKINAT 135
 QY 505 GGLPSAAGGVA--VGEPPVAMGNSGGQGTTPRVPGRVVALGQTV-QASDSLTGAET 561
 DB 136 GGLPTIPINARRVPHIGDVLAINCP-----YNLQITQITQIISATG-RIG 180
 QY 562 LN-----GLTQFAAIQPGSGPVNGLGVVGMNT 593
 DB 181 LNPTGRQNFLOTASINHGSGALVNSLGLMGINT 217

RESULT 15

ID DEGP CHLPN STANDARD; PRT; 488 AA.
 AC Q926T0; Q9JQD7; Q9K1W4;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable serine protease do-like precursor (EC 3.4.21.-).
 GN DEGP OR HTRA OR CPN0979 OR CP0877 OR CPB1016.
 OS Chlamydia pneumoniae (Chlamydia phila pneumoniae).
 CC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia phila.
 OK NCBI_TaxID=83558;
 [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=CWL029;
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Olinger J., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389(1999).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935; Gill S.R., Heidelberg J.F.,
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Utterback T., Berry K., Bass S.,
 RA White O., Hickey E.K., Peterson J., Dodson R.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 RT pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and whole genome sequences of Chlamydia pneumoniae J138
 RL Nucleic Acids Res. 28:2311-2314(2000).
 [4]
 RN SEQUENCE FROM N.A.
 RP STRAIN=TW-183;
 RA Gang M.M., Schumacher A., Muehldorfer I., Bensch K.W., Schaefer K.P.,
 RA Schneider S., Pohl T., Essig A., Marre R., Melchers K.;
 RT "The genome sequence of Chlamydia pneumoniae TW183 and comparison with
 RT other Chlamydia strains based on whole genome sequence analysis.";
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: Belongs to peptidase family S2C.
 CC -!- SIMILARITY: Contains 2 PDZ/DHR domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AE001678; AAD19116.1; -
 DR EMBL; AE002246; AAF38665.1; -
 DR EMBL; AP002548; BAA99186.1; -
 DR EMBL; AE017160; NAP98945.1; ALT_INIT.
 DR PIR; G72011; G72011.
 DR PIR; H86612; H86612.
 DR MEROPS; S01.273; -
 DR PHCI-2DPAGE; Q926T0; -
 DR TIGR; CP0877; -
 DR InterPro; IPR009003; Cys_Ser_trypsin.
 DR InterPro; IPR001478; PDZ_Protein_S1.
 DR InterPro; IPR001254; Peptidase_S1.
 DR InterPro; IPR001940; Peptidase_S1C.
 DR Pfam; PF00595; PDZ; 2.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS0106; PDZ; 2.
 KW Hydrolase; Serine protease; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 488
 FT DOMAIN 119 280
 FT DOMAIN 281 372
 FT DOMAIN 388 476
 FT ACT_SITE 134 134
 FT ACT_SITE 164 164
 FT ACT_SITE 238 238
 FT CONFLICT 17 17
 SQ SEQUENCE 488 AA; 52311 MW; 0EE7E0F88F106F49 CRC64;
 Query Match 7.5%; Score 225; DB 1; Length 488;
 Best Local Similarity 28.6%; Pred. No. 9.9e-05;
 Matches 83; Conservative 42; Mismatches 99; Indels 66; Gaps 14;
 QY 330 QAWAANQAVTPAARALPLTSLTSAERGPQMLG-----GLPVGQMGARAGGSLGV 382
 DB 7 RSLAV-LVGSSLLALPLS-----GQAVGKESRVSELPQDVLKLEISGGFSKV 54
 QY 383 -----LRVPRPVMPHSPAAGDIAPPALSDQRFAD-----FPALPLDPSANVAQ 427
 DB 55 ATKATPAVVYTESPFKQAVTH-ESPGRRGYENPFDFNDEFNRRFFGLPSQ-----R 107
 QY 428 VGPVWVINTKLYNNVAGTGVIPDNGVLTNNHVIAGATDINAFSVSGQTYGVVDV 487
 DB 108 EKPQ-----SKNAVGTGFLVSPDGYIVTNNHVEDTGKIHV-TLHDGQKYPATV 156
 QY 488 VGYDRTQDVAVLQRLGAGGLPSAAGG--VAVGEPPVAMGNSGGQGTTPRVPGRVVAL 545
 DB 157 IGLDPKTDLAVIKIK-SQNPYLSFGNSDHLKVGDMWAIAGNPFGLQAT--VTVGVISAK 213
 QY 546 GQT-VQASDSLTGAETLNGLIQFADALQPGDSGPPVNGLGQVVGNTA 594
 DB 214 GRNQHLHAD-----FEDFIQTDAAINPGNSGGLINIDGQVIGNTA 255

Search completed: June 30, 2004, 16:49:45
 Job time : 10.7035 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:43:31 ; Search time 15.5767 Seconds
(without alignments)
3680.509 Million cell updates/sec

Title: US-09-597-796C-10

Perfect score: 3007

Sequence: 1 HMMHHHHMVFGLPPEIN.....SGGPFVNGLGQVGMNTAAS 596

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_78:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1949	64.8	391	2 B70608	probable PPE prote
2	1656.5	55.1	396	2 H70741	probable PPE prote
3	1583	52.6	393	2 C70568	probable PPE prote
4	989	32.9	355	2 F70983	probable serine pr
5	779.5	25.9	393	2 G70929	probable PPE prote
6	778.5	25.9	361	2 S47170	hypothetical prote
7	754.5	25.1	354	2 A87242	probable secreted
8	753.5	25.1	409	2 A70932	probable PPE prote
9	737.5	24.5	403	2 H70931	probable PPE prote
10	737	24.5	423	2 B70931	probable PPE prote
11	731	24.3	421	2 H87056	PPE-family protein
12	705	23.4	408	2 G70925	probable PPE prote
13	702	23.3	391	2 B70625	probable PPE prote
14	702	23.3	413	2 F70560	probable PPE prote
15	690	22.9	468	2 B70932	probable PPE prote
16	688.5	22.9	463	2 C70931	probable PPE prote
17	675	22.4	380	2 A70646	probable PPE prote
18	668.5	22.2	394	2 G70881	probable PPE prote
19	667.5	22.2	385	2 H70503	probable PPE prote
20	635	21.1	350	2 H70929	probable PPE prote
21	633.5	21.1	365	2 H70929	probable PPE prote
22	615.5	20.5	402	2 A70882	probable PPE prote
23	606.5	20.2	423	2 C70582	probable PPE prote
24	602.5	20.0	391	2 D70922	probable PPE prote
25	595	19.8	394	2 A70504	probable PPE prote
26	589	19.6	406	2 F70675	probable PPE prote
27	575.5	19.1	391	2 A70663	probable PPE prote
28	521.5	17.3	3300	2 D70575	probable PPE prote
29	477.5	15.9	3716	2 B70969	probable PPE prote

30	473	15.7	180	2	G70834	probable PPE prote
31	467	15.5	655	2	A70931	probable PPE prote
32	462.5	15.4	580	2	G70570	probable PPE prote
33	458	15.2	3157	2	B70969	probable PPE prote
34	450	15.0	2523	2	F70846	probable PPE prote
35	449	14.9	963	2	B70524	probable PPE prote
36	448	14.9	678	2	A70762	probable PPE prote
37	447.5	14.9	582	2	F70675	probable PPE prote
38	444	14.8	346	2	H70874	probable PPE prote
39	442	14.7	487	2	C70830	probable PPE prote
40	441	14.7	479	2	D70676	probable PPE prote
41	441	14.7	552	2	D70604	probable PPE prote
42	437.5	14.5	615	2	E70663	probable PPE prote
43	429.5	14.3	645	2	F70925	probable PPE prote
44	428.5	14.3	1053	2	B70987	probable PPE prote
45	426.5	14.2	443	2	C70780	probable PPE prote

ALIGNMENTS

RESULT 1

B70608

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C/Species: Mycobacterium tuberculosis

C/Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C/Accession: B70608

R/Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Rajandream, M.A.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Connor, R.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A/Reference number: A70500; MUID:98295987; PMID:9634230

A/Accession: B70608

A/Status: Preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-391 <COL>

A/Cross-references: GB:Z93777; GB:AL123456; NID:G3261726; PIDN:CAB07839.1; PID:G311073; I

A/Experimental source: strain H37RV

C/Genetics:

A/Gene: PPE

Query Match	64.8%	Score	1949;	DB	2;	Length	391;
Best Local Similarity	100.0%	Pred. No.	1.3e-95;				
Matches	391;	Conservative	0;	Mismatches	0;	Indels	0;
Gaps	0;						
Qy	9	MVDFGALPPEINSARMVAGPGSASLVAAACQMWDSVASDLFSAASAFQSVVWGLTVGSWIG	68				
Db	1	MVDFGALPPEINSARMVAGPGSASLVAAACQMWDSVASDLFSAASAFQSVVWGLTVGSWIG	60				
Qy	69	SSAGLVAAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI	128				
Db	61	SSAGLVAAASPYVAMWSVTAGQAEITAAQVRVAAAAYETAYGLTVPPVIAENRAELMI	120				
Qy	129	LIAITNLGQNTPAIAVNEAYGEMWAQDAAMFGVYAAATATATATALLPEEPAPMTSAG	188				
Db	121	LIAITNLGQNTPAIAVNEAYGEMWAQDAAMFGVYAAATATATATALLPEEPAPMTSAG	180				
Qy	189	LLEQAAVBEASPTAAANQLMNNVPAALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI	248				
Db	181	LLEQAAVBEASPTAAANQLMNNVPAALQQLAQPTQGTTPSSKLGGLWKTVPSPHRSPI	240				
Qy	249	MVSMANNHSMNTNSGVSMNTLLSSMLKGFAPAPAAAQVOTAAQNGVRAMSSLGSSG	308				
Db	241	MVSMANNHSMNTNSGVSMNTLLSSMLKGFAPAPAAAQVOTAAQNGVRAMSSLGSSG	300				
Qy	309	LGSGVAANI-GRASVGSLSVPOAWAANCAVTPAAEALPLTISLTAABERGPGOMLGLPV	368				
Db	301	LGSGVAANI-GRASVGSLSVPOAWAANCAVTPAAEALPLTISLTAABERGPGOMLGLPV	360				
Qy	369	GQMGAPAGGLSGVLRVPPRPVPHSPAAG	399				

Db 361 GQMGARAGGSLGVLVRPPRPYVMPHSPAAG 391

RESULT 2

H70741

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: H70741

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID: 98295987; PMID: 9634230

A: Accession: H70741

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-396 <COL>

A: Cross-references: GB:275555; GB:AL123456; NID: g3261608; PIDN: CAA99966.1; PID: e250360;

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: PPE

Query Match 55.1%; Score 1656.5; DB 2; Length 396;

Best Local Similarity 85.1%; Pred. No. 3.1e-80;

Matches 336; Conservative 18; Mismatches 34; Indels 7; Gaps 3;

QY 9 MVDFGALPPPEINARMYAGPGSASLVAAQWMDVSADLFSAAAFQSVVWGLTVGSGWIG 68

Db 1 MVDFGALPPPEINARMYAGPGSASLVAAQWMDVSADLFSAAAFQSVVWGLTVGSGWIG 60

QY 69 SSAGLMVAASPYYVWMSVTAQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128

Db 61 SSAGLMVAASPYYVWMSVTAQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 120

QY 129 LIATNLGQNTPALAVNEAEYGEWQAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 188

Db 121 LIATNLGQNTPALAVNEAEYGEWQAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 180

QY 189 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTGCTTSSKGLGWLKTVSPHRSPIIN 248

Db 181 LLEQAAVEEADITAAANQLMNNVPQALQQAQPTGCTTSSKGLGWLKTVSPHRSPIIN 240

QY 249 MVSNNHNSMTNSGVSMNTLSSMLKGFAPAAAQAVOTAQNGVRAMSS---LGSSL 304

Db 241 IVSMLNHNHNSMTNSGVSMNTLSSMLKGFAPAAAQAVOTAQNGVRAMSS---LGSSL 299

QY 305 GSSGLGGVAAANLGRAASVGSLSVPPQAWAANAQAVTPAARALPLTSLTSAARGPQOMLG 364

Db 300 GSSGLGGVAAANLGRAASVGSLSVPPQAWAANAQAVTPAARALPLTSLTSAARGPQOMLG 359

QY 365 GLPVGMQARAG--GGLSGVLVRPPRPYVMPHSPAAG 399

Db 360 GLPLGQLTNSGGGFGVSNALRMPPRAYVMPRVEAAG 396

RESULT 3

C70568

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: C70568

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID: 98295987; PMID: 9634230

A: Accession: C70568

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-393 <COL>

A: Cross-references: GB:295390; GB:AL123456; NID: g3261766; PIDN: CAB08702.1; PID: e316074; I

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: PPE

Query Match 52.6%; Score 1583; DB 2; Length 393;

Best Local Similarity 81.5%; Pred. No. 2.3e-76;

Matches 322; Conservative 20; Mismatches 47; Indels 6; Gaps 3;

QY 9 MVDFGALPPPEINARMYAGPGSASLVAAQWMDVSADLFSAAAFQSVVWGLTVGSGWIG 68

Db 1 MVDFGALPPPEINARMYAGPGSASLVAAQWMDVSADLFSAAAFQSVVWGLTVGSGWIG 60

QY 69 SSAGLMVAASPYYVWMSVTAQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128

Db 61 SSAGLMVAASPYYVWMSVTAQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMT 120

QY 129 LIATNLGQNTPALAVNEAEYGEWQAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 188

Db 121 LIATNLGQNTPALAVNEAEYGEWQAQDAAMFGYAAATATATATLLPPEAPEMTSAGG 180

QY 189 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTGCTTSSKGLGWLKTVSPHRSPIIN 248

Db 181 LLEQAAVEEADITAAANQLMNNVPQALQQAQPTGCTTSSKGLGWLKTVSPHRSPIIN 240

QY 249 MVSNNHNSMTNSGVSMNTLSSMLKGFAPAAAQAVOTAQNGVRAMSS---LGSSL 304

Db 241 VSSANNHNSMTNSGVSMNTLSSMLKGFAPAAAQAVOTAQNGVRAMSS---LGSSL 299

QY 305 GSSGLGGVAAANLGRAASVGSLSVPPQAWAANAQAVTPAARALPLTSLTSAARGPQOMLG 364

Db 300 GSSGLGGVAAANLGRAASVGSLSVPPQAWAANAQAVTPAARALPLTSLTSAARGPQOMLG 359

QY 365 GLPVGMQARAGGGLSGVLVRPPRPYVMPHSPAAG 399

Db 360 GLPLGHL-SVNAGSGINNALRVPARAYAIPTPAAG 393

RESULT 4

F70983

probable serine proteinase pepA - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 09-Dec-2002

C:Accession: F70983

R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID: 98295987; PMID: 9634230

A: Accession: F70983

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-355 <COL>

A: Cross-references: GB:296071; GB:AL123456; NID: g3242254; PIDN: CAB09453.1; PID: g2181967

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: pepA

C: Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; tryp

Query Match 32.9%; Score 989; DB 2; Length 355;

Best Local Similarity 98.0%; Pred. No. 3.4e-45;

Matches 196; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 397 AAGDIAPALSSQDFADFPALPLDPSAMVAQGVQVNVNINTKLGYNNAVAGTGIVIDPN 456

Db 28 APAQAAAPALSSQDFADFPALPLDPSAMVAQGVQVNVNINTKLGYNNAVAGTGIVIDPN 87

QY 457 GVLITNNHVIAGATDINAFSVGSGQTGVGVGVDRTQDVAVIOLRGAGGLPSAIGGV 516

Db 88 GVVLTNNHVIAGATDINAFSGSGQTYGVVVGVYDRTQDVAVLQVRGAGGLPSAAIGGGV 147
QY 517 AVGEPVVMAGNSGGGGGTTPRVPGRVVALGQTQVQASDSLTGAETTLNGLIQFDAAIQPD 576
Db 148 AVGEPVVMAGNSGGGGGTTPRVPGRVVALGQTQVQASDSLTGAETTLNGLIQFDAAIQPD 207
QY 577 SGGPVVNGLGQVVGWNTAAS 596
Db 208 SGGPVVNGLGQVVGWNTAAS 227

RESULT 5
G70929
probable PPE protein - Mycobacterium tuberculosis (strain H37Rv)
C:Species: Mycobacterium tuberculosis
C>Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70929
R:Coile, S.T.; Garnier, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70929
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-393 <COL>
A:Cross-references: GB:AL02021; GB:AL123456; NID:G3250699; PIDN:CAAL17711.1; PID:el25460
A:Experimental source: strain H37Rv
C:Genetics:
A:Gene: PPE

Query Match 25.9%; Score 779.5; DB 2; Length 393;
Best Local Similarity 43.7%; Pred. No. 4e-34;
Matches 179; Conservative 65; Mismatches 129; Indels 37; Gaps 10;

QY 10 VDFGLPPBINGARMYAGPSSASLVAAQMDVSASDLFSAASAPQSVVWGLTVGWSIGS 69
Db 1 MDFGLPPEVNSVRYVYAGPSAPMYAASAWKGLAELSSAATGYETVITQLSSSGWLP 60

QY 70 SAGLVAAASPVVAMSVTAGAELTAAQVRVAAAYETAYGLTVPVPPVIAENRAELMIL 129
Db 61 AASAAAEAVPVVAMSVAAQAQEAATQARAAAFAFAAFAATVPPLIAANRASLMQL 120

QY 130 IATNLGQNTPALAVNEAGYGBWMAQDAAMAGYAAATATATATLLPPEEAPEMTSAGGL 189
Db 121 ISTNVFGQNTSALAAEAQYGEWMAQDSAMTAYAGSSASASA-VTPSTPPQIANPAAQ 179

QY 190 LEQAAAVEASDTAAA--NQLMNNVPQALQQLAQPTQGTTPSSKLGGLWK----- 237
Db 180 GTQAAAVATAAGTAQSTLTEMITGLPNAQLSLTSLPQQSS-NGPLSLWLQILFGTPNPT 238

QY 238 -----TVSPHRSPISNMVSANNHMTNSGVSNVTLSSMLKGFAPAAAQAVQTAA 290
Db 239 STSALTLDIOPVASFYNTGFLPYFSIGNGNFIQSAKTL-GLIGSAPAAVA-----AA 292

QY 291 QNGVRAMSSLGSSSLGSGGVAANLGRAASVGSLSVPQAWAA-ANOAVTPAARALPLT 349
Db 293 GDAKGLPGLGWLG---GGPVAGLGNAAVSKLSVPPVWSGPLPGSVTPGAAPLFS 348

QY 350 SUTSAAERGGQMLGPLGVQMGARAGGLSGVLPRPVPYMPHSPPAG 399
Db 349 TVSAAPEAAPGLGLGLPL---AGAGGAGAGP-RYGFEPYTMARPPFPAG 393

RESULT 6
S47170
hypothetical protein 34K - Mycobacterium paratuberculosis
C:Species: Mycobacterium paratuberculosis
C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Dec-2002
C:Accession: S47170
R:Cameron, R.M.; Stevenson, K.; Inglis, N.F.; Klausen, J.; Sharp, J.M.

submitted to the EMBL Data Library, June 1993
A:Description: Isolation and characterisation of a 34kDa protein of Mycobacterium paratub
A:Reference number: S47170
A:Accession: S47170
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-361 <CAM>
A:Cross-references: EMBL:Z23092; NID:G505550; PIDN:CAA80638.1; PID:G505551
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; trypsin

Query Match 25.9%; Score 778.5; DB 2; Length 361;
Best Local Similarity 66.0%; Pred. No. 4.1e-34;
Matches 155; Conservative 28; Mismatches 37; Indels 15; Gaps 2;

QY 362 MLGGLPVQMGARAGGLSGVLPRPVPYMPHSPPAGDIAPPALSDRADPPALPLDP 421
Db 15 LVGVLTVVGLGLGSGVG-----LAPASA--APSGLAIDRFADPLADP 59

QY 422 SAMVAQVGPVWNTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQ 481
Db 60 SAMVQVGPVWNTKFGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQ 119

QY 482 TYGVVGVYDRTQDVAVLQVRGAGGLPSAAIGGGVAVGEPVVMAGNSGGGGTTPRVPGR 541
Db 120 TYAVDVGVYDRTQDVAVLQVRGAGGLPTATIGGEATVGEPIVALGNVGGQGTTPNAVAGK 179

QY 542 VVALGQTVQASDSLTGAETTLNGLIQFDAAIQPDGSGPVVWGLGQVVGWNTAAS 596
Db 180 VVALNQSVSATLTLTGAENLGLLIQADAFIKPDGSGPVWNSAGQVIGVDTRAT 234

RESULT 7
A87242
Probable secreted serine proteinase [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C>Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Dec-2002
C:Accession: A87242
R:Coile, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Hor
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: A87242
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-354 <STO>
A:Cross-references: GB:AL450380; NID:gl3093863; PIDN:CAC32191.1; GSPDB:GN00147
A:Gene: ML2659
C:Genetics:
C:Superfamily: Escherichia coli trypsin-like proteinase degs; GLGF domain homology; trypsin

Query Match 25.18%; Score 754.5; DB 2; Length 354;
Best Local Similarity 73.3%; Pred. No. 7.3e-33;
Matches 151; Conservative 25; Mismatches 25; Indels 5; Gaps 2;

QY 391 VMHSPAAIGDIAPPALSDRFPALPLDPFSAMVAQVGPVWNTKLGYNNAVGAGTG 450
Db 27 VVPGS--ATPSGPSTLALDRFSNRFPPLNPAWVA---PQVWNISRLGVNSAVGAGTG 81

QY 451 IVIDPNGVLTNNHVIAGATDINAFSGSGQTVGVYDRTQDVAVLQVRGAGGLPSA 510
Db 82 IVIDSSGVLTTNNHVIAGATDISAFDVGNGKTVGVYDRTQDVAVLQVRGAGGLPSA 141

QY 511 AIGGGVAVGEPVVMAGNSGGGGTTPRVPGRVVALGQTQVQASDSLTGAETTLNGLIQFDA 570
Db 142 VIGGDVAIGEPVVALGNTGGGLPSVLPGRVVALGQTQVQASSEPLTGAQETLSGLIQVDA 201

QY 571 AIQPDGSGPVVWGLGQVVGWNTAAS 596
Db 202 PIKPDGSGPVVWNSRGQVVGWNTAAT 227

RESULT 8

A70932 probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: A70932

R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID: 9825987; PMID: 9634230

A: Accession: A70932

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-409 <COL>

A: Cross-references: GB:AL022021; GB:AL123456; NID: g3250699; PIDN: CAA17729.1; PID: e125461

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: PPE

Query Match 25.1%; Score 753.5; DB 2; Length 409;

Best Local Similarity 42.6%; Pred. No. 9.8e-33;

Matches 182; Conservative 54; Mismatches 152; Indels 39; Gaps 11;

QY 10 VDFGALPPEINSARMYAGPGSASIVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWS 69

Db 1 MDFGALPPEINSARMYAGPGSGGLLAADAWDAELYSAAASYGSTEGLTVAPWGP 60

QY 70 SAGLMVAAASPYVAMSVTGAQBELTAAQVRVAAAAYETAYGLTVPPVTAENRAELMIL 129

Db 61 SSITMAAAVAPYVAMISVTGAQEAQAKIAAGVYETAFAATVPVPPVTEANRALLMSL 120

QY 130 IATNLGONTPTAVNEAEYGEWMAQDAAMFGVAAATATATATATLLPFEAPMTSAGGL 189

Db 121 VATNIFGONTPTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 179

QY 190 LEQAAVVEEASDTAAA-----NQLMNNVPQALQQLAQTQGTTPSSKLGGLWKTVPSPH 242

Db 180 AAQSAVVAQAAGAAASDDITAQLSQLISLPLTQSLA--TTATATASAG--WDIV--- 232

QY 243 RSPISNMVSMNMHMTNSGVS-----MTNTLSSMLKGFAPAAAQAVQTAQNGVRAM 297

Db 233 LQSITTLTANLTPYISIGLALPGGWLTFFGQILGLAQNAPGVAAALLGPKAAAGALSPL 292

QY 298 SSL-GSSLGS-SGLGGVAAANLGRAASVGLSVFQAWAANQAVTPAARALPLTSLTSA- 354

Db 293 APLGGYIGDITPLGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 352

QY 355 AERPGQMLGLPVQMGARAGGGL-----SGVLVPPPPVYVPHSPAGDIAPPALSD 409

Db 353 AARAPGALFGEMALSSLAGALAGTAVRSGAGAARV-----AGGSVTEVDASTT 401

QY 410 RFADFFA 416

Db 402 TIIVIPA 408

RESULT 9

H70931

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: H70931

R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID: 9825987; PMID: 9634230

A: Accession: H70931

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-403 <COL>

A: Cross-references: GB:AL022021; GB:AL123456; NID: g3250699; PIDN: CAA17728.1; PID: e125461

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: PPE

Query Match 24.5%; Score 737.5; DB 2; Length 403;

Best Local Similarity 42.6%; Pred. No. 6.7e-32;

Matches 181; Conservative 60; Mismatches 123; Indels 61; Gaps 12;

QY 10 VDFGALPPEINSARMYAGPGSASIVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWS 69

Db 5 LDFATLPEINSARMYSGAGSAPVLAASAASAHGSLASALSYSVSLTLTGEWHGP 64

QY 70 SAGLMVAAASPYVAMSVTGAQBELTAAQVRVAAAAYETAYGLTVPPVTAENRAELMIL 129

Db 65 ASASMTAAAPYVAMSVTAVRAEQAGQAQAEAAAAYEAAFAATVPVPPVTEANRAQLMAL 124

QY 130 IATNLGONTPTAVNEAEYGEWMAQDAAMFGVAAATATATATATLLPFEAPMTSAGGL 189

Db 125 IATNLGONTPTAVNEAEYGEWMAQDAAMFGVAAATATATATATLLPFEAPMTSAGGL 183

QY 190 LEQAAVVEEASDTAAA-----NQLMNNVPQALQQLAQTQGT-----TPSS 230

Db 184 AAQSAATTAATGASAGAQOTTLTSLIAAIPSVLQGLSSSTAATFASGPGSLGIVGSGSS 243

QY 231 KLGGLWKTVPSPHSPISNMVSMNMHMTNSGYSM-TNTLS---SMLKGFAPAAAQAV 286

Db 244 WLDKLMALDLPN-----SNFWNTIASGGLFPLSNTIAPPLGLGGVAAADAAGDV 293

QY 287 QTAQNGVRAMSSSLGSSL-----GSSGLGGVAAANLGRAASVGLSVFQAWAANQAVTP 341

Db 294 LGATSG-----GLGALVAPLGSAGGLGTVAGLGNAAATVGLTSPSWTAAAPLSP 348

QY 342 AARAL---PLTSLTSAERGPQMLGGLPVQMGARAGGGLSGVLVPP-----RPYMPH 394

Db 349 LGSALGCTPMVAPPVPAAG---MPGPFMTMGQGGFG-----RAVPOYGFPRNFVAR 398

QY 395 SPAAG 399

Db 399 PPAAG 403

RESULT 10

B70931

probable PPE protein - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: B70931

R: Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A: Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A: Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A: Reference number: A70500; MUID: 9825987; PMID: 9634230

A: Accession: B70931

A: Status: preliminary; nucleic acid sequence not shown; translation not shown

A: Molecule type: DNA

A: Residues: 1-423 <COL>

A: Cross-references: GB:AL022021; GB:AL123456; NID: g3250699; PIDN: CAA17722.1; PID: e125461

A: Experimental source: strain H37RV

C: Genetics:

A: Gene: PPE

Query Match 24.5%; Score 737; DB 2; Length 423;

Best Local Similarity 41.8%; Pred. No. 7.6e-32;

Matches 183; Conservative 50; Mismatches 125; Indels 80; Gaps 12;

QY 10 VDFGALPPEINSARMYAGPGSASIVAAQWMDSVASDLFSAASAFQSVVWGLTVGSGWS 69

Db 1 MDFGLPPPEINSGRMYTGPFGPMLAAATWDLGAVELHATAAGVASELSALT-GANSGP 59
QY 70 SAGLWVAASPYVAMSVTGAQAEHTAAQVRVAAAAYETAYGLTVPPPPVIAENRAELMIL 129
Db 60 SSTSMASAAAPYVAMSVATAVHAELAGACARLALAAAYEAFAATVPPPPVIAANRAQLAVL 119
QY 130 IATNLLGONTPTAIVNEAYGEMWQAQDAAMCYAAATATATATLPPPEAPEMTSAGL 189
Db 120 IATNIFGONTPTAIVNEAYGEMWQAQDAAMCYAAATATATATLPPPEAPEMTSAGL 178
QY 190 LEQAAAVEASDTAAAN-----OLMNVPOALQLOAQT-----QGTFP-----S 229
Db 179 GQSSAVAGTAATAAGGNLQSFQPLL SAVPALQGLALPTASQASATPQWTDLGNLS 238
QY 230 SKLGGWLKTVSPHRSPISNMVSVANNHMTNSGVSMNTLSSMLKGFAPAAAQVQTA 289
Db 239 TELGG-AVTGPYTFP-----GVLPPSGVPYLLIGQSVL-----V 271
QY 290 AONGVRAVMSGLS-----SIGSSGLGGG-VAANLGRAASVGSLS 327
Db 272 TONGQGVSAALLKIGKGPITGALAPLAEFALHTPPIILGSEGLGGVSASGIGRAGLVGKLS 331
QY 328 VPQWAAANQAVTPAARALPLTSLTS---AABERGPGQMLGGLPVGQMGARAGGSLGVLR 384
Db 332 VPQGWTVAAPEIPSPAAALQATRLAAAPTAAATDAGALLGGWALSGLAGRAAGSTG--- 388
QY 385 VPPRPVPMHSPAAAGDIA 402
Db 389 ---HPTGSAAPAVAGAAA 403

RESULT 11
H87056
PPE-family protein [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: H87056
R: Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Holt,
R.; Davies, R.M.; Devlin, K.; Duchoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Squ
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: A86909; MUID:21128732; PMID:11234002
A:Accession: H87056
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-421 <STO>
A:Cross-references: GB:AL450380; NID:gl3093150; PIDN:CAC31563.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML1182

Query Match 24.3%; Score 731; DB 2; Length 421;
Best Local Similarity 41.1%; Pred. No. 1.6e-31;
Matches 174; Conservative 55; Mismatches 160; Indels 34; Gaps 8;
QY 9 MVDGALPPEINSGRMYTGPFGPMLAAATWDLGAVELHATAAGVASELSALT-GANSGP 59
Db 1 MFDFAALSPEINSGRMYTGPFGPMLAAATWDLGAVELHATAAGVASELSALT-GANSGP 59
QY 69 SAGLWVAASPYVAMSVTGAQAEHTAAQVRVAAAAYETAYGLTVPPPPVIAENRAELMIL 128
Db 60 ESAAALAEVTPYKWLTONAASAEHTATQTLVAANAYETATMTVPPPLMVFVNRQAQL 119
QY 129 LIATNLLGONTPTAIVNEAYGEMWQAQDAAMCYAAATATATLPPPEAPEMTSAGL 188
Db 120 LINSNTFGQNTAIAENAEYEMWQAQDAAMCYAAATATATLPPPEAPEMTSAGL 179
QY 189 L-----LEQAAAVEASDTAAANQMLNN-----VPOALQ-----LAOP 222
Db 180 LAQVEEVVEVVEVVEVVEVVEVVEVVEVVEVVEVVEVVEVVEVVEVVEVVEVVEV 239

QY 223 TQGTTPSSKLG--LWKTVPSPHRSPISNMVSVANNHMTNSGVSMNTLSSMLKGFAPA 280
Db 240 QTAVPDSSSSAAAAPQLWGGFAQLHLSPITLSTMTNNHAGMANAGLSVNGSGSAMKSLAP- 298
QY 281 AAAQAVOTAAQNGVRAMSSLSGSSGLGGVAAANLGRAASVGSLSVPOWAAANQAVT 340
Db 299 TTTTAAESAFAKMSAVQSTGRGLGSSSGHVTVAQLGRAASIGSLRVPQTWTTASQPT 358
QY 341 PAARALPLTSLTSAERPGQML-GGLPVGQM---GARAGGSLGVLRVPPRPVPMHSP 396
Db 359 AATRALSPARVAVATESAPLGGGLPMAPVPGGSGTGGVNTALRLQFRAFVMPRNP 418
QY 397 AAG 399
Db 419 AAG 421

RESULT 12
G70925
Probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: G70925
R: Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.;
Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987; PMID:9634230
A:Accession: G70925
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-408 <COL>
A:Cross-references: GB:Z74024; GB:AL123456; NID:g3250700; PIDN:CAA98377.1; PID:e1301025;
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: PPE

Query Match 23.4%; Score 705; DB 2; Length 408;
Best Local Similarity 41.8%; Pred. No. 3.5e-30;
Matches 182; Conservative 47; Mismatches 134; Indels 72; Gaps 14;
QY 10 VDFGALPPEINSGRMYTGPFGPMLAAATWDLGAVELHATAAGVASELSALT-GANSGP 69
Db 1 MDFGVLPEINSGRMYTGPFGPMLAAATWDLGAVELHATAAGVASELSALT-GANSGP 60
QY 70 SAGLWVAASPYVAMSVTGAQAEHTAAQVRVAAAAYETAYGLTVPPPPVIAENRAELMIL 129
Db 61 AAASMWAAVTPYVAVLSATAGAQAGQARAAAAAYELAFAMTVPPPVVAVNALLVAL 120
QY 130 IATNLLGONTPTAIVNEAYGEMWQAQDAAMCYAAATATATLPPPEAPEMTSAGL 189
Db 121 VATNFFGONTPTAIVNEAYGEMWQAQDAAMCYAAATATLPPPEAPEMTSAGL 179
QY 190 LEQAAA-----VEASDTAAANQMLNN-----VP--QALQ--LAQQTGTTPSSKL--- 232
Db 180 AGQAAATVSTVPPPLATTAAPVQLQSLSTLIPWYSAALQWLAENLLGLTPNRRITV 239
QY 233 -----GGLWKTVPSPHRSPISNMVSVANNHMTNSGVSMNTLSSMLKGFAPAAA 283
Db 240 RLIGISVDEGL-----LOFEASLAQQAIPGTPGGAG--DSGSSVLDSHGPIFA 287
QY 284 QAVQTAAQNGVRAMSSLS--GSSLSG-----SGLGGVAAANLGRAASVGSLS 327
Db 288 -----GPRASPVAGGAGGVQTPQPYWYVWALDRESIGGSVAALKGSSAGLS 338
QY 328 VPQWAAANQAVTPAARALP---LTSLSAERPGQMLGGLPVGQMGARAGGSLGVLR 384
Db 339 VPPDWAARAWANPAANWLLPGDDVTALRGTAENA---LLRGFPVASAGQSTGGGF---VHK 393
QY 385 VPPRPVPMHSPAAAG 399

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:52:58 ; Search time 46.4747 Seconds
(without alignments)
3625.462 Million cell updates/sec

Title: US-09-597-796C-10

Perfect score: 3007
Sequence: 1 HHHHHHHHVDGALPPEIN.....SGGPVYVGLGVGMNTAAS 596

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1166195 seqs, 282705291 residues

Total number of hits satisfying chosen parameters: 1166195

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/prodata/2/pubpaa/US06_PCT_NEW_PUB.pep.*
- 3: /cgn2_6/prodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/prodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/prodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/prodata/2/pubpaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/prodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/prodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/prodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/prodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/prodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/prodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/prodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/prodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/prodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/prodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/prodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/prodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3007	100.0	596	9	US-09-287-849-26
2	3007	100.0	596	12	US-09-886-349A-20
3	3007	100.0	596	14	US-10-359-460-26
4	3007	100.0	596	14	US-10-098-732A-20
5	2946	98.0	729	15	US-10-369-983-21
6	2946	98.0	813	15	US-10-369-983-15
7	2946	98.0	825	15	US-10-369-983-14
8	2946	98.0	875	15	US-10-369-983-13
9	2946	98.0	930	14	US-10-098-732A-65
10	2946	98.0	930	15	US-10-369-983-12
11	2946	98.0	1016	15	US-10-369-983-18
12	2946	98.0	1022	15	US-10-369-983-17
13	2946	98.0	1154	15	US-10-369-983-16
14	2943	97.9	723	12	US-09-886-349A-18
15	2943	97.9	723	14	US-10-098-732A-18

16	2943	97.9	729	15	US-10-369-983-22	Sequence 22, Appl
17	2941	97.8	729	9	US-09-287-849-2	Sequence 2, Appl
18	2941	97.8	729	12	US-09-886-349A-16	Sequence 16, Appl
19	2941	97.8	729	14	US-10-359-460-2	Sequence 2, Appl
20	2941	97.8	723	14	US-10-098-732A-16	Sequence 16, Appl
21	2941	97.8	729	15	US-10-359-459-2	Sequence 2, Appl
22	2033.5	67.6	600	9	US-09-287-849-22	Sequence 22, Appl
23	2033.5	67.6	600	14	US-10-359-460-22	Sequence 4, Appl
24	1955	65.0	1010	15	US-10-369-983-4	Sequence 14, Appl
25	1949	64.8	391	12	US-09-886-349A-14	Sequence 102, App
26	1949	64.8	391	14	US-10-193-002-102	Sequence 107, App
27	1949	64.8	391	14	US-10-084-843-107	Sequence 14, Appl
28	1949	64.8	391	14	US-10-098-732A-14	Sequence 14, Appl
29	1949	64.8	723	15	US-10-369-983-2	Sequence 8, Appl
30	1934	64.3	391	12	US-09-872-186-8	Sequence 106, App
31	1652.5	55.0	396	14	US-10-193-002-106	Sequence 111, App
32	1652.5	55.0	396	14	US-10-084-843-111	Sequence 62455, A
33	1583	52.6	393	12	US-10-282-122A-62455	Sequence 64892, A
34	1583	52.6	393	12	US-10-282-122A-64892	Sequence 104, App
35	1486.5	49.4	359	14	US-10-193-002-104	Sequence 109, App
36	1486.5	49.4	359	14	US-10-084-843-109	Sequence 8, Appl
37	1189	39.5	358	9	US-09-287-849-8	Sequence 8, Appl
38	1189	39.5	358	14	US-10-359-460-8	Sequence 12, Appl
39	1187	39.5	263	12	US-09-886-349A-12	Sequence 92, Appl
40	1187	39.5	263	14	US-10-193-002-92	Sequence 91, Appl
41	1187	39.5	263	14	US-10-084-843-91	Sequence 12, Appl
42	1187	39.5	263	14	US-10-098-732A-12	Sequence 161, App
43	989	32.9	355	9	US-09-712-363-161	Sequence 4, Appl
44	987	32.8	330	12	US-09-886-349A-4	Sequence 4, Appl
45	987	32.8	330	14	US-10-098-732A-4	

ALIGNMENTS

RESULT 1

US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. US2002009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match 100.0%; Score 3007; DB 9; Length 596;
Best Local Similarity 100.0%; Pred. No. 5e-202;

Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HMGHHHHMVFGLPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWG 60
DB 1 HMGHHHHMVFGLPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWG 60

QY 61 LTVGSWIGSSAGLWVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIA 120
DB 61 LTVGSWIGSSAGLWVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIA 120

QY 121 ENRAELMLIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATILLPEEA 180
DB 121 ENRAELMLIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATILLPEEA 180

QY 181 PEMSAGLLBQAAAVEASDTAAANQLMNNVPQALQOQAOPTGTTTPSSKLGGLWKTYS 240
DB 181 PEMSAGLLBQAAAVEASDTAAANQLMNNVPQALQOQAOPTGTTTPSSKLGGLWKTYS 240

QY 241 PHRSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAONGVRAMSSL 300
DB 241 PHRSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAONGVRAMSSL 300

QY 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAABERGPG 360
DB 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAABERGPG 360

QY 361 QMLGGLPVQMGARAGGGLSVLRVPPRPYVMPHSPAAAGDIAPALSDQDFADFPALPLD 420
DB 361 QMLGGLPVQMGARAGGGLSVLRVPPRPYVMPHSPAAAGDIAPALSDQDFADFPALPLD 420

QY 421 PSAMVAQVGPVQVNTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480
DB 421 PSAMVAQVGPVQVNTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480

QY 481 QTYGVVGVYDRTQDVAVLQRGAGGLPSAAIGGGVAVGEPVVMGNSGGGGTTPRAVPG 540
DB 481 QTYGVVGVYDRTQDVAVLQRGAGGLPSAAIGGGVAVGEPVVMGNSGGGGTTPRAVPG 540

QY 541 RVVALGQTVQASDSLTGAETLNGLIQFDDAAIQFGDSGGPVVNGLGQVVGWNTAAS 596
DB 541 RVVALGQTVQASDSLTGAETLNGLIQFDDAAIQFGDSGGPVVNGLGQVVGWNTAAS 596

RESULT 2
US-09-886-349A-20
; Sequence 20, Application US/09886349A
; Publication No. US20040086523A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Reed, Steven
; APPLICANT: Alderson, Mark
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: US/09/886,349A
; CURRENT APPLICATION NUMBER: 2001-06-20
; PRIOR FILING DATE: 2001-06-20
; PRIOR FILING DATE: 2000-06-20
; PRIOR FILING DATE: 2000-06-20
; PRIOR FILING DATE: 2001-02-01
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: bi-fusion
; OTHER INFORMATION: protein TBH9-Ra35 (Designated MTB59F)
US-09-886-349A-20

Query Match 100.0%; Score 3007; DB 12; Length 596;
Best Local Similarity 100.0%; Pred. No. 5e-202;

Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HMGHHHHMVFGLPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWG 60
DB 1 HMGHHHHMVFGLPPEINSARMYAGPGSASLVAAQWMDSVASDLFSAASAFQSVVWG 60

QY 61 LTVGSWIGSSAGLWVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIA 120
DB 61 LTVGSWIGSSAGLWVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPVIA 120

QY 121 ENRAELMLIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATILLPEEA 180
DB 121 ENRAELMLIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATILLPEEA 180

QY 181 PEMSAGLLBQAAAVEASDTAAANQLMNNVPQALQOQAOPTGTTTPSSKLGGLWKTYS 240
DB 181 PEMSAGLLBQAAAVEASDTAAANQLMNNVPQALQOQAOPTGTTTPSSKLGGLWKTYS 240

QY 241 PHRSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAONGVRAMSSL 300
DB 241 PHRSPISNMVMANNHMTNSGVSMNTLSSMLKGFAPAAAQAQVTAQAONGVRAMSSL 300

QY 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAABERGPG 360
DB 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAABERGPG 360

QY 361 QMLGGLPVQMGARAGGGLSVLRVPPRPYVMPHSPAAAGDIAPALSDQDFADFPALPLD 420
DB 361 QMLGGLPVQMGARAGGGLSVLRVPPRPYVMPHSPAAAGDIAPALSDQDFADFPALPLD 420

QY 421 PSAMVAQVGPVQVNTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480
DB 421 PSAMVAQVGPVQVNTKLGYNNAVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSG 480

QY 481 QTYGVVGVYDRTQDVAVLQRGAGGLPSAAIGGGVAVGEPVVMGNSGGGGTTPRAVPG 540
DB 481 QTYGVVGVYDRTQDVAVLQRGAGGLPSAAIGGGVAVGEPVVMGNSGGGGTTPRAVPG 540

QY 541 RVVALGQTVQASDSLTGAETLNGLIQFDDAAIQFGDSGGPVVNGLGQVVGWNTAAS 596
DB 541 RVVALGQTVQASDSLTGAETLNGLIQFDDAAIQFGDSGGPVVNGLGQVVGWNTAAS 596

RESULT 3
US-10-359-460-26
; Sequence 26, Application US/10359460
; Publication No. US20030147911A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/10/359,460
; CURRENT FILING DATE: 2003-02-05
; PRIOR FILING DATE: 2003-02-05
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1999-04-07
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-03-13
; PRIOR FILING DATE: 1997-10-01
; PRIOR FILING DATE: 1999-02-18
; PRIOR FILING DATE: 1998-02-18
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-04-07
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1


```
;
;
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:MTB72F
; US-10-369-983-21
;
; Query Match          98.0%; Score 2946; DB 15; Length 729;
; Best Local Similarity 100.0%; Pred. No. 1.2e-197;
; Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 9 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 68
; DB 142 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 201
;
; QY 69 SSAGLVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128
; DB 202 SSAGLVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 261
;
; QY 129 LIATNLLGQNTPAIAVNEAAYGEMWQDAAAMFGYAAATATATATLLPFEAPEMTSAGG 188
; DB 262 LIATNLLGQNTPAIAVNEAAYGEMWQDAAAMFGYAAATATATATLLPFEAPEMTSAGG 321
;
; QY 189 LLEQAAAVEASDTAAANQLMNNVPOALQOLAQTOGTTPSSKLGGLWKTVPSPHSPISN 248
; DB 322 LLEQAAAVEASDTAAANQLMNNVPOALQOLAQTOGTTPSSKLGGLWKTVPSPHSPISN 381
;
; QY 249 MVSMAHHMNTSGVSMNTTLLSSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSLGSSG 308
; DB 382 MVSMAHHMNTSGVSMNTTLLSSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSLGSSG 441
;
; QY 309 LGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 368
; DB 442 LGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 501
;
; QY 369 GQMGARAGGGLSGVLVPPPPYVMPHSPAAAGDIAPPALSDQDFADPPALPLDPSAMVAQV 428
; DB 502 GQMGARAGGGLSGVLVPPPPYVMPHSPAAAGDIAPPALSDQDFADPPALPLDPSAMVAQV 561
;
; QY 429 GPQVNTIKLYNNVAVGAGTGIVDPNGVVLNNHVIAGATDINAFSVGSGQTYGVDVV 488
; DB 562 GPQVNTIKLYNNVAVGAGTGIVDPNGVVLNNHVIAGATDINAFSVGSGQTYGVDVV 621
;
; QY 489 GYDRTQDVAVLQIRGAGGLPSAIIGGVAVGEPVVMGNSGGGGTPPRAVPGRVVAGQT 548
; DB 622 GYDRTQDVAVLQIRGAGGLPSAIIGGVAVGEPVVMGNSGGGGTPPRAVPGRVVAGQT 681
;
; QY 549 VQASDSLTGAETLNGLIQFDAAIQPGDSGPPVNGLGQVVGMMNTAAS 596
; DB 682 VQASDSLTGAETLNGLIQFDAAIQPGDSGPPVNGLGQVVGMMNTAAS 729
;
; RESULT 6
; US-10-369-983-15
; Sequence 15, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 813
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:fusion protein
; OTHER INFORMATION: MTB81F (MTB72F-DPV)
; US-10-369-983-14
;
; Query Match          98.0%; Score 2946; DB 15; Length 813;
; Best Local Similarity 100.0%; Pred. No. 1.4e-197;
; Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
; QY 9 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 68
; DB 142 MVDGALPPPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSWIG 201
;
; QY 69 SSAGLVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128
; DB 202 SSAGLVAAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 261
;
; QY 129 LIATNLLGQNTPAIAVNEAAYGEMWQDAAAMFGYAAATATATATLLPFEAPEMTSAGG 188
; DB 262 LIATNLLGQNTPAIAVNEAAYGEMWQDAAAMFGYAAATATATATLLPFEAPEMTSAGG 321
;
; QY 189 LLEQAAAVEASDTAAANQLMNNVPOALQOLAQTOGTTPSSKLGGLWKTVPSPHSPISN 248
; DB 322 LLEQAAAVEASDTAAANQLMNNVPOALQOLAQTOGTTPSSKLGGLWKTVPSPHSPISN 381
;
; QY 249 MVSMAHHMNTSGVSMNTTLLSSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSLGSSG 308
; DB 382 MVSMAHHMNTSGVSMNTTLLSSMLKGFAPAAAQAQVTAQNGVRAMSSLGSSLGSSG 441
;
; QY 309 LGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 368
; DB 442 LGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGPQMLGGLPV 501
;
; QY 369 GQMGARAGGGLSGVLVPPPPYVMPHSPAAAGDIAPPALSDQDFADPPALPLDPSAMVAQV 428
; DB 502 GQMGARAGGGLSGVLVPPPPYVMPHSPAAAGDIAPPALSDQDFADPPALPLDPSAMVAQV 561
;
; QY 429 GPQVNTIKLYNNVAVGAGTGIVDPNGVVLNNHVIAGATDINAFSVGSGQTYGVDVV 488
; DB 562 GPQVNTIKLYNNVAVGAGTGIVDPNGVVLNNHVIAGATDINAFSVGSGQTYGVDVV 621
;
; QY 489 GYDRTQDVAVLQIRGAGGLPSAIIGGVAVGEPVVMGNSGGGGTPPRAVPGRVVAGQT 548
; DB 622 GYDRTQDVAVLQIRGAGGLPSAIIGGVAVGEPVVMGNSGGGGTPPRAVPGRVVAGQT 681
;
; QY 549 VQASDSLTGAETLNGLIQFDAAIQPGDSGPPVNGLGQVVGMMNTAAS 596
; DB 682 VQASDSLTGAETLNGLIQFDAAIQPGDSGPPVNGLGQVVGMMNTAAS 729
;
; RESULT 7
; US-10-369-983-14
; Sequence 14, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 825
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:fusion protein
; OTHER INFORMATION: MTB83F (MTB72F-MTI)
; US-10-369-983-14
```

Query Match 98.0%; Score 2946; DB 15; Length 825;
Best Local Similarity 100.0%; Pred. No. 1.4e-137;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MYDFGALPEINSARMYAGPGSASLVAAQAQMDSDVASDLFSAASAFQSVVWGLTVGSWIG 68
DB 142 MYDFGALPEINSARMYAGPGSASLVAAQAQMDSDVASDLFSAASAFQSVVWGLTVGSWIG 201

QY 69 SSAGLWAAASPYVWMSVTAGQAELETAQVRVAAAAYETAYGLTPPVIAENRAELMI 128
DB 202 SSAGLWAAASPYVWMSVTAGQAELETAQVRVAAAAYETAYGLTPPVIAENRAELMI 261

QY 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
DB 262 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 321

QY 189 LLEQAAAEEASDTAAANQLMNNVPOALQOLAQPTGCTTPSSKLGGLWKTVPSPHSPISN 248
DB 322 LLEQAAAEEASDTAAANQLMNNVPOALQOLAQPTGCTTPSSKLGGLWKTVPSPHSPISN 381

QY 249 MYSMANNHMTNSGVSMTNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLSGSSG 308
DB 382 MYSMANNHMTNSGVSMTNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLSGSSG 441

QY 309 LGGVAAANLGRAASVGSLSVPOQAAAAQAVQTAQNGVRAMSSLSGSSG 368
DB 442 LGGVAAANLGRAASVGSLSVPOQAAAAQAVQTAQNGVRAMSSLSGSSG 501

QY 369 GQMGARAGGGLSGVLRVPRPYVMPHSPAAQDIAPPALSDQRFADFPALPLDPSAMVAQV 428
DB 502 GQMGARAGGGLSGVLRVPRPYVMPHSPAAQDIAPPALSDQRFADFPALPLDPSAMVAQV 561

QY 429 GPQVYNINTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDDV 488
DB 562 GPQVYNINTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDDV 621

QY 489 GYDRTQDVAVLQRLGAGGLPSAAIGGVAVGEPVWAMGNSGGQGTTPRAVGRVVALGQT 548
DB 622 GYDRTQDVAVLQRLGAGGLPSAAIGGVAVGEPVWAMGNSGGQGTTPRAVGRVVALGQT 681

QY 549 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVWGLGVVGMNTAAS 596
DB 682 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVWGLGVVGMNTAAS 729

RESULT 8
US-10-369-983-13
; Sequence 13, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Gudarian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 875
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB89F (MTB72F-Erd14)
US-10-369-983-13

Query Match 98.0%; Score 2946; DB 15; Length 875;
Best Local Similarity 100.0%; Pred. No. 1.5e-137;

Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MYDFGALPEINSARMYAGPGSASLVAAQAQMDSDVASDLFSAASAFQSVVWGLTVGSWIG 68
DB 142 MYDFGALPEINSARMYAGPGSASLVAAQAQMDSDVASDLFSAASAFQSVVWGLTVGSWIG 201

QY 69 SSAGLWAAASPYVWMSVTAGQAELETAQVRVAAAAYETAYGLTPPVIAENRAELMI 128
DB 202 SSAGLWAAASPYVWMSVTAGQAELETAQVRVAAAAYETAYGLTPPVIAENRAELMI 261

QY 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
DB 262 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 321

QY 189 LLEQAAAEEASDTAAANQLMNNVPOALQOLAQPTGCTTPSSKLGGLWKTVPSPHSPISN 248
DB 322 LLEQAAAEEASDTAAANQLMNNVPOALQOLAQPTGCTTPSSKLGGLWKTVPSPHSPISN 381

QY 249 MYSMANNHMTNSGVSMTNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLSGSSG 308
DB 382 MYSMANNHMTNSGVSMTNTLSSMLKGFAPAAAAQAVQTAQNGVRAMSSLSGSSG 441

QY 309 LGGVAAANLGRAASVGSLSVPOQAAAAQAVQTAQNGVRAMSSLSGSSG 368
DB 442 LGGVAAANLGRAASVGSLSVPOQAAAAQAVQTAQNGVRAMSSLSGSSG 501

QY 369 GQMGARAGGGLSGVLRVPRPYVMPHSPAAQDIAPPALSDQRFADFPALPLDPSAMVAQV 428
DB 502 GQMGARAGGGLSGVLRVPRPYVMPHSPAAQDIAPPALSDQRFADFPALPLDPSAMVAQV 561

QY 429 GPQVYNINTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDDV 488
DB 562 GPQVYNINTKLGYNNAVGAGTGIVDPNGVLTNNHVIAGATDINAFSGSGQTYGVDDV 621

QY 489 GYDRTQDVAVLQRLGAGGLPSAAIGGVAVGEPVWAMGNSGGQGTTPRAVGRVVALGQT 548
DB 622 GYDRTQDVAVLQRLGAGGLPSAAIGGVAVGEPVWAMGNSGGQGTTPRAVGRVVALGQT 681

QY 549 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVWGLGVVGMNTAAS 596
DB 682 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVWGLGVVGMNTAAS 729

RESULT 9
US-10-098-732A-65
; Sequence 65, Application US/10098732A
; Publication No. US20030175294A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Brannon, Mark
; APPLICANT: Gudarian, Jeffrey
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
; TITLE OF INVENTION: Leishmania Antigen
; FILE REFERENCE: 014058-012010US
; CURRENT APPLICATION NUMBER: US/10/098,732A
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/275,837
; PRIOR FILING DATE: 2001-03-13
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 65
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: MTB72F-MAPS
; OTHER INFORMATION: (r95f) fusion construct, TB MTB72F (Rai12-TbH9-Ra35)
; OTHER INFORMATION: linked to Leishmania thiol-specific antioxidant
; OTHER INFORMATION: (TSA or MAPS)
US-10-098-732A-65

Query Match 98.0%; Score 2946; DB 14; Length 930;

Best Local Similarity 100.0%; Pred. No. 1.6e-197;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MVDGALPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSMIG 68
DB 142 MVDGALPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSMIG 201
QY 69 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128
DB 202 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 261
QY 129 LIATNLLGONTPALAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 188
DB 262 LIATNLLGONTPALAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 321
QY 189 LLEQAAAVEEASDTAAANQLMNVPQALQLOAQTOQTTPSSKLGGLWKTVPSPHRSPISN 248
DB 322 LLEQAAAVEEASDTAAANQLMNVPQALQLOAQTOQTTPSSKLGGLWKTVPSPHRSPISN 381
QY 249 MVSMAHNMHMTNSGVSMVTNLTSSMLKGFAPAAAAQAVQTAQNGVAMSSLSGLSSG 308
DB 382 MVSMAHNMHMTNSGVSMVTNLTSSMLKGFAPAAAAQAVQTAQNGVAMSSLSGLSSG 441
QY 309 LGGVAAANLGRAASVGLSVPOQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 368
DB 442 LGGVAAANLGRAASVGLSVPOQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 501
QY 369 GQMGARAGGGLSVLRVPPRYPMHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 428
DB 502 GQMGARAGGGLSVLRVPPRYPMHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 561
QY 429 GPQVNTNKLGYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVV 488
DB 562 GPQVNTNKLGYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVV 621
QY 489 GYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEPVAVMGNSGGGTPPRAVPGRVVALGQT 548
DB 622 GYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEPVAVMGNSGGGTPPRAVPGRVVALGQT 681
QY 549 VQASDSLTGAETLNGLIQFDAAIQDSDGSGPVVNGLGQVVGNTAAS 596
DB 682 VQASDSLTGAETLNGLIQFDAAIQDSDGSGPVVNGLGQVVGNTAAS 729

RESULT 10
US-10-369-983-12
; Sequence 12, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; PRIOR FILING DATE: 2003-02-18
; PRIOR FILING DATE: 2003-02-18
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: R95F (MTB72F-NAPs)
US-10-369-983-12

Query Match 98.0%; Score 2946; DB 15; Length 930;
Best Local Similarity 100.0%; Pred. No. 1.6e-197;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MVDGALPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSMIG 68
DB 142 MVDGALPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSMIG 201
QY 69 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 128
DB 202 SSAGLMVAASPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPPIAENRAELMI 261
QY 129 LIATNLLGONTPALAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 188
DB 262 LIATNLLGONTPALAVNEAEYGEWMAQDAAMFGYAAATATATATLPPFEAPMTSAGG 321
QY 189 LLEQAAAVEEASDTAAANQLMNVPQALQLOAQTOQTTPSSKLGGLWKTVPSPHRSPISN 248
DB 322 LLEQAAAVEEASDTAAANQLMNVPQALQLOAQTOQTTPSSKLGGLWKTVPSPHRSPISN 381
QY 249 MVSMAHNMHMTNSGVSMVTNLTSSMLKGFAPAAAAQAVQTAQNGVAMSSLSGLSSG 308
DB 382 MVSMAHNMHMTNSGVSMVTNLTSSMLKGFAPAAAAQAVQTAQNGVAMSSLSGLSSG 441
QY 309 LGGVAAANLGRAASVGLSVPOQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 368
DB 442 LGGVAAANLGRAASVGLSVPOQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 501
QY 369 GQMGARAGGGLSVLRVPPRYPMHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 428
DB 502 GQMGARAGGGLSVLRVPPRYPMHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 561
QY 429 GPQVNTNKLGYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVV 488
DB 562 GPQVNTNKLGYNNVAGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDVV 621
QY 489 GYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEPVAVMGNSGGGTPPRAVPGRVVALGQT 548
DB 622 GYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEPVAVMGNSGGGTPPRAVPGRVVALGQT 681
QY 549 VQASDSLTGAETLNGLIQFDAAIQDSDGSGPVVNGLGQVVGNTAAS 596
DB 682 VQASDSLTGAETLNGLIQFDAAIQDSDGSGPVVNGLGQVVGNTAAS 729

RESULT 11
US-10-369-983-18
; Sequence 18, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR FILING DATE: 2003-02-18
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1016
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; OTHER INFORMATION: MTB103F (MTB72F-85b)
US-10-369-983-18

Query Match 98.0%; Score 2946; DB 15; Length 1016;
Best Local Similarity 100.0%; Pred. No. 1.8e-197;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 MVDGALPPEINARMYAGPGSASLVAAQWDSVADLFSAAAFQSVVWGLTVGSMIG 68

Db	142	MDVFGALPEIINISARIYAGPGSASLVAAQMWDSVADLFSAAFAQSVVWGLTVGSWIG	201
Qy	69	SSAGLMAAASPVVAWMSVTAGCAELTAAQVRAAAAYETAYGLTVPPPIAENRAELMI	128
Db	202	SSAGLMAAASPVVAWMSVTAGCAELTAAQVRAAAAYETAYGLTVPPPIAENRAELMI	261
Qy	129	LIATNLLGONTPAIAVNEAEYGEWMAODAAAMFGYAAATATATATALLPFEAPEMTSAG	188
Db	262	LIATNLLGONTPAIAVNEAEYGEWMAODAAAMFGYAAATATATATALLPFEAPEMTSAG	321
Qy	189	LLEQAAVEEAESDTAAANQLMNNVPOALQOQLAQTQGTTPSSKLGGLUKTVSPHRSPIN	248
Db	322	LLEQAAVEEAESDTAAANQLMNNVPOALQOQLAQTQGTTPSSKLGGLUKTVSPHRSPIN	381
Qy	249	MYSMANNHSMWNTSGVSMNTLSSMLKGFAPAAAQAVQTAAQNGVRAMSSLGSSLGSSG	308
Db	382	MYSMANNHSMWNTSGVSMNTLSSMLKGFAPAAAQAVQTAAQNGVRAMSSLGSSLGSSG	441
Qy	309	LGCGVAAANLGRAASVGSLSVQAAANAQAVTPAARALPLTSTSAERCPGOMLGLBV	368
Db	442	LGCGVAAANLGRAASVGSLSVQAAANAQAVTPAARALPLTSTSAERCPGOMLGLBV	501
Qy	369	GQMGARAGGSLGVLVPPRPVYMPHSPAAGDIAAPPALSQDRPADFPALPLDPSAMVAQV	428
Db	502	GQMGARAGGSLGVLVPPRPVYMPHSPAAGDIAAPPALSQDRPADFPALPLDPSAMVAQV	561
Qy	429	GPOVNIINTKLGYNNAVAGAGTGIVDPNGVVLTNHVIIAGATINAFSVSGGQTYGVGV	488
Db	562	GPOVNIINTKLGYNNAVAGAGTGIVDPNGVVLTNHVIIAGATINAFSVSGGQTYGVGV	621
Qy	489	GYDRTQDVAVLQLRGAGGLPSAAIIGGVAVGEPVVAVMGNSGGGGTTPRAVGRVVALGOT	548
Db	622	GYDRTQDVAVLQLRGAGGLPSAAIIGGVAVGEPVVAVMGNSGGGGTTPRAVGRVVALGOT	681
Qy	549	VOASDSLTCASSETLNGLIQFDAAIQPGDSGGPVVNGLGQVGMNTAAS	596
Db	682	VOASDSLTCASSETLNGLIQFDAAIQPGDSGGPVVNGLGQVGMNTAAS	729

```

RESULT 12
US-10-369-983-17
; Sequence 17, Application US/10369983
; Publication No. US20030235593A1
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Guderian, Jeff
; APPLICANT: Reed, Steven
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
; FILE REFERENCE: 014058-009081US
; CURRENT APPLICATION NUMBER: US/10/369,983
; CURRENT FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/357,351
; PRIOR FILING DATE: 2002-02-15
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 17
; LENGTH: 1022
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: fusion protein
; - OTHER INFORMATION: MTB102tm2F (MTB102FTM2, MTB72F-hTCC#1)
US-10-369-983-17

```

Qy	69	SSAGLWAAASPYVAMSVTAGAQLTAAQVVRVAAAAYETAYGLTVPVPTAENRAELMI	128
Db	202	SSAGLWAAASPYVAMSVTAGAQLTAAQVVRVAAAAYETAYGLTVPVPTAENRAELMI	261
Qy	129	LIATNLGQNTPAIAVNEABYGENWAOADAAAFGYAAATATATATLPPFEEAPEMTSAGG	188
Db	262	LIATNLGQNTPAIAVNEABYGENWAOADAAAFGYAAATATATATLPPFEEAPEMTSAGG	321
Qy	189	LLFQAAAVEASDTAAANQLMNNVPQALQLOAQPTQGTTPSSKLGGIWKTVSPHRSPISN	248
Db	322	LLFQAAAVEASDTAAANQLMNNVPQALQLOAQPTQGTTPSSKLGGIWKTVSPHRSPISN	381
Qy	249	MVSMANNHSMTNSGYSMTNTLSLSMLKGFAPAAAAQAVQTAAQNGVRAMSSLSGSSLAGSG	308
Db	382	MVSMANNHSMTNSGYSMTNTLSLSMLKGFAPAAAAQAVQTAAQNGVRAMSSLSGSSLAGSG	441
Qy	309	LGCGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERPGQMLGGLPV	368
Db	442	LGCGVAANLGRAASVGSLSVPQAWAANAQVTPAARALPLTSLTSAERPGQMLGGLPV	501
Qy	369	GQNGARAGGSLGVLRVPRPYVWPHSPAAAGDIAPPALSDRPFADFPALPLDPSAMVAQV	428
Db	502	GQNGARAGGSLGVLRVPRPYVWPHSPAAAGDIAPPALSDRPFADFPALPLDPSAMVAQV	561
Qy	429	GPQVWINTKLGYNNAVGAGTGVIDPQNGVVLNNHVIAGATDINAFSVGSGQTYGVDDV	488
Db	562	GPQVWINTKLGYNNAVGAGTGVIDPQNGVVLNNHVIAGATDINAFSVGSGQTYGVDDV	621
Qy	489	GYDRTQDVAVLQRGAGGLPSAAIGGSVAVGEPPVAMGNSGGGGTTPRVPGRVVALGQT	548
Db	622	GYDRTQDVAVLQRGAGGLPSAAIGGSVAVGEPPVAMGNSGGGGTTPRVPGRVVALGQT	681
Qy	549	VQASDSLTAEEETLNGLIQFDAAIQPCDGGPPVNGIQGVVGMNTAAS	596
Db	682	VQASDSLTAEEETLNGLIQFDAAIQPCDGGPPVNGIQGVVGMNTAAS	729

```

RESULT 13
US-10-369-983-16
/ Sequence 16, Application US/10369983
/ Publication No. US20030235593A1
/ GENERAL INFORMATION:
/ APPLICANT: Skeiky, Yasir
/ APPLICANT: Guderian, Jeff
/ APPLICANT: Reed, Steven
/ APPLICANT: Corixa Corporation
/ TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
/ FILE REFERENCE: 014058-009081US
/ CURRENT APPLICATION NUMBER: US/10/369,983
/ CURRENT FILING DATE: 2003-02-18
/ PRIOR APPLICATION NUMBER: 2003-02-18
/ PRIOR APPLICATION NUMBER: US 60/357,351
/ PRIOR FILING DATE: 2002-02-15
/ NUMBER OF SEQ ID NOS: 22
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 16
/ LENGTH: 1154
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial
/ OTHER INFORMATION: MTB114F (MTB72F-mTCC#2)
US-10-369-983-16

```

	Query Match	98.0%;	Score 2946;	DB 15;	Length 1154;
	Best Local Similarity	100.0%;	Pred. NO. 2.2e-197;		
	Matches 588;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	9	WVDFGALPPEINSARMYAPGSGASVLAQAQMWDVSASDLFSAASAPQSVVWGLTVGSWIG	68		
Db	142	WVDFGALPPEINSARMYAPGSGASVLAQAQMWDVSASDLFSAASAPQSVVWGLTVGSWIG	201		
QY	69	SSAGLMLVAAASPYVAMWSVTACGAEILTAQOVRAVAAAAYETAXGLTVPPVPIAENRAELMI	138		

202 SSAGLWVAASPYVAMSVTAGQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 261
 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPEMTSAGG 188
 262 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPEMTSAGG 321
 189 LLEQAAAVEASDTAAANQMLMNVPAALQOLAQTOGTTPSSKLGGLWKTVPSPHSPISN 248
 322 LLEQAAAVEASDTAAANQMLMNVPAALQOLAQTOGTTPSSKLGGLWKTVPSPHSPISN 381
 249 MVSMAHNMHMTNSGVSMTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 308
 382 MVSMAHNMHMTNSGVSMTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 441
 309 LGGGVAANLGRAASVGSLSVPPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 368
 442 LGGGVAANLGRAASVGSLSVPPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 501
 369 GQMGARAGGGLSGVLRVPPPYVMPHSPAAQDIAPPALSOQDRFADFPALPLDPSAMVAQV 428
 502 GQMGARAGGGLSGVLRVPPPYVMPHSPAAQDIAPPALSOQDRFADFPALPLDPSAMVAQV 561
 429 GPQVNVINTKLGYNNAVAGTGIVDPNGVVLNNHVIAGATDINAFSVGSGGQTYGVVV 488
 562 GPQVNVINTKLGYNNAVAGTGIVDPNGVVLNNHVIAGATDINAFSVGSGGQTYGVVV 621
 489 GYDRTQDVAVLQLRGAGLPSAAIQPGDGGPVVNGLGQVVGWMTAAS 596
 622 GYDRTQDVAVLQLRGAGLPSAAIQPGDGGPVVNGLGQVVGWMTAAS 729

RESULT 14
 US-09-886-349A-18
 ; Sequence 18, Application US/09886349A
 ; Publication No. US20040086523A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Reed, Steven
 ; APPLICANT: Alderson, Mark
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Fusion Proteins of Mycobacterium Tuberculosis
 ; FILE REFERENCE: 014058-009070US
 ; CURRENT APPLICATION NUMBER: US/09/886,349A
 ; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: US 09/597,796
 ; PRIOR FILING DATE: 2000-06-20
 ; PRIOR APPLICATION NUMBER: US 60/265,737
 ; PRIOR FILING DATE: 2001-02-01
 ; NUMBER OF SEQ ID NOS: 50
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 18
 ; LENGTH: 729
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:MTB72FmutSA
 ; OTHER INFORMATION: (Ra12-TBHp-Ra35MutSA)
 US-09-886-349A-18

Query Match 97.9%; Score 2943; DB 12; Length 729;
 Best Local Similarity 99.8%; Pred. No. 2e-197;
 Matches 587; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 9 MVDGALPPEINSMARMYAGPGSASLVAAQWDSVASDLFSAASAFOSVVMGLTVGWSIG 68
 142 MVDGALPPEINSMARMYAGPGSASLVAAQWDSVASDLFSAASAFOSVVMGLTVGWSIG 201
 69 SSAGLWVAASPYVAMSVTAGQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 128

202 SSAGLWVAASPYVAMSVTAGQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 261
 129 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPEMTSAGG 188
 262 LIATNLLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPPEEAPEMTSAGG 321
 189 LLEQAAAVEASDTAAANQMLMNVPAALQOLAQTOGTTPSSKLGGLWKTVPSPHSPISN 248
 322 LLEQAAAVEASDTAAANQMLMNVPAALQOLAQTOGTTPSSKLGGLWKTVPSPHSPISN 381
 249 MVSMAHNMHMTNSGVSMTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 308
 382 MVSMAHNMHMTNSGVSMTNTLSMLKGFAPAAAQAQVTAQAQNGVRAMSSLGSSLGSSG 441
 309 LGGGVAANLGRAASVGSLSVPPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 368
 442 LGGGVAANLGRAASVGSLSVPPQAWAANAQVTPAARALPLTSLTSAERGPQMLGGLPV 501
 369 GQMGARAGGGLSGVLRVPPPYVMPHSPAAQDIAPPALSOQDRFADFPALPLDPSAMVAQV 428
 502 GQMGARAGGGLSGVLRVPPPYVMPHSPAAQDIAPPALSOQDRFADFPALPLDPSAMVAQV 561
 429 GPQVNVINTKLGYNNAVAGTGIVDPNGVVLNNHVIAGATDINAFSVGSGGQTYGVVV 488
 562 GPQVNVINTKLGYNNAVAGTGIVDPNGVVLNNHVIAGATDINAFSVGSGGQTYGVVV 621
 489 GYDRTQDVAVLQLRGAGLPSAAIQPGDGGPVVNGLGQVVGWMTAAS 596
 622 GYDRTQDVAVLQLRGAGLPSAAIQPGDGGPVVNGLGQVVGWMTAAS 729

RESULT 15
 US-10-098-732A-18
 ; Sequence 18, Application US/10098732A
 ; Publication No. US20030175294A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Skeiky, Yasir
 ; APPLICANT: Brannon, Mark
 ; APPLICANT: Guderian, Jeffrey
 ; APPLICANT: Corixa Corporation
 ; TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
 ; FILE REFERENCE: 014058-012010US
 ; CURRENT APPLICATION NUMBER: US/10/098,732A
 ; CURRENT FILING DATE: 2003-04-29
 ; PRIOR APPLICATION NUMBER: US 60/275,837
 ; PRIOR FILING DATE: 2001-03-13
 ; NUMBER OF SEQ ID NOS: 80
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 18
 ; LENGTH: 729
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence:MTB72FmutSA
 ; OTHER INFORMATION: (Ra12-TBHp-Ra35MutSA)
 US-10-098-732A-18

Query Match 97.9%; Score 2943; DB 14; Length 729;
 Best Local Similarity 99.8%; Pred. No. 2e-197;
 Matches 587; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 9 MVDGALPPEINSMARMYAGPGSASLVAAQWDSVASDLFSAASAFOSVVMGLTVGWSIG 68
 142 MVDGALPPEINSMARMYAGPGSASLVAAQWDSVASDLFSAASAFOSVVMGLTVGWSIG 201
 69 SSAGLWVAASPYVAMSVTAGQELTAAQVRVAAAAYETAYGLTVPVPIAENRAELMI 128


```
Db      202  SSAGLMVAASPVTAMSVTAGAEELTAAQVRAAAAYETAYGLTVPPPEVIAENRAELMI 261
QY      129  LIATNLLGQNTPAIAVNEAEYGEEMQAQDAAMFGVAAAATATATATILLPEEPAPENTSAGG 188
Db      262  LIATNLLGQNTPAIAVNEAEYGEEMQAQDAAMFGVAAAATATATATILLPEEPAPENTSAGG 321
QY      189  LLEQAAAVEEASDSTAANQIMNNVPOALQOLAQPTQGITPSSKLGGLWKTVPSPHRSPISN 248
Db      322  LLEQAAAVEEASDSTAANQIMNNVPOALQOLAQPTQGITPSSKLGGLWKTVPSPHRSPISN 381
QY      249  MYSMANNHSMNTNSGVSMNTNTLSSMLKGFAPAPAAAQAVOTAAQNGVYRAMSSILGSSLGSSG 308
Db      382  MYSMANNHSMNTNSGVSMNTNTLSSMLKGFAPAPAAAQAVOTAAQNGVYRAMSSILGSSLGSSG 441
QY      309  LGGGVAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 368
Db      442  LGGGVAANLGRAASVGSLSVPOAWAAANQAVTPAARALPLTSLTSAABERGPGQMLGGLPV 501
QY      369  GQMGARAGGGLSGVLRVPPRYMPHSPAAAGDIAPPALSCDEFADFPALPLDPSAMVAQV 428
Db      502  GQMGARAGGGLSGVLRVPPRYMPHSPAAAGDIAPPALSCDEFADFPALPLDPSAMVAQV 561
QY      429  GPQVNVNINTKLYNNNAVAGTGVIVDPNGVILTNHVIAGATDINAFVSGSQTYGVVDV 488
Db      562  GPQVNVNINTKLYNNNAVAGTGVIVDPNGVILTNHVIAGATDINAFVSGSQTYGVVDV 621
QY      489  GYDRTQDVAVLQIRGAGGLPSAIIAGGVAVGEPVYAMGNSGGQGTTPRAVPCRVALQQT 548
Db      622  GYDRTQDVAVLQIRGAGGLPSAIIAGGVAVGEPVYAMGNSGGQGTTPRAVPCRVALQQT 681
QY      549  VQASDSLTAETLNGLIQFDAAIQFSDSGPVPVNGLGQVVGMTAAS 596
Db      682  VQASDSLTAETLNGLIQFDAAIQFSDSGPVPVNGLGQVVGMTAAS 729
```

Search completed: June 30, 2004, 17:14:46
Job time : 48.4747 secs

1. The first part of the document discusses the importance of maintaining accurate records of all transactions and the role of the accounting department in ensuring the integrity of the financial statements. It also highlights the need for regular audits and the importance of transparency in financial reporting.

2. The second part of the document focuses on the implementation of internal controls to prevent fraud and ensure the accuracy of financial data. It outlines the key components of a robust internal control system, including segregation of duties, authorization procedures, and regular monitoring and evaluation.

3. The third part of the document addresses the challenges faced by organizations in managing their financial resources effectively. It discusses the importance of budgeting, forecasting, and cost management, and provides practical advice on how to overcome common financial management challenges.

4. The fourth part of the document explores the role of technology in modern accounting and finance. It discusses the benefits of using accounting software and the importance of staying up-to-date with the latest technological advancements in the field.

5. The fifth part of the document concludes by emphasizing the importance of ethical behavior in accounting and finance. It discusses the role of accountants and financial managers in upholding the highest standards of ethical conduct and the consequences of unethical behavior.

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:44:57 ; Search time 18.1302 Seconds
(without alignments)
1697.113 Million cell updates/sec

Title: US-09-597-796C-10
Perfect score: 3007
Sequence: 1 HMEHHHHMVDFFGALPPEIN.....SGGPVNVGLQGVGMNTAAS 596

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA.*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.psp.*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.psp.*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.psp.*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.psp.*
5: /cgn2_6/ptodata/2/iaa/PCRTUS_COMB.psp.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.psp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3007	100.0	596	4	US-09-287-849-26
2	2341	97.8	729	4	US-09-223-040-2
3	2341	97.8	729	4	US-09-287-849-2
4	2033.5	67.6	600	4	US-09-287-849-22
5	1949	64.8	391	3	US-08-818-112-107
6	1949	64.8	391	4	US-08-818-111-102
7	1949	64.8	391	4	US-09-056-556-107
8	1949	64.8	391	4	US-09-072-967-102
9	1949	64.8	391	4	US-09-072-967-107
10	1652.5	55.0	396	3	US-08-818-112-111
11	1652.5	55.0	396	4	US-08-818-111-106
12	1652.5	55.0	396	4	US-09-056-556-111
13	1652.5	55.0	396	4	US-09-072-967-106
14	1652.5	55.0	396	4	US-09-072-967-109
15	1486.5	49.4	359	3	US-08-818-112-109
16	1486.5	49.4	359	4	US-08-818-111-104
17	1486.5	49.4	359	4	US-09-056-556-109
18	1486.5	49.4	359	4	US-09-072-967-104
19	1486.5	49.4	359	4	US-09-072-967-109
20	1187	39.5	358	4	US-09-287-849-8
21	1187	39.5	263	3	US-08-818-112-91
22	1187	39.5	263	4	US-08-818-111-92
23	1187	39.5	263	4	US-09-056-556-91
24	1187	39.5	263	4	US-09-072-967-92
25	1187	39.5	263	4	US-09-072-967-91
26	983	32.7	355	3	US-08-818-112-79
27	983	32.7	355	4	US-08-818-111-80

28	983	32.7	355	4	US-09-056-556-79	Sequence 79, Appl
29	983	32.7	355	4	US-09-072-967-80	Sequence 80, Appl
30	983	32.7	355	4	US-09-072-967-79	Sequence 79, Appl
31	770.5	25.6	400	4	US-09-073-009-126	Sequence 126, App
32	770.5	25.6	400	4	US-09-073-010-126	Sequence 126, App
33	607.5	20.2	710	4	US-09-287-849-16	Sequence 16, Appl
34	607.5	20.2	856	4	US-09-287-849-12	Sequence 12, Appl
35	606.5	20.2	423	4	US-09-073-009-142	Sequence 142, App
36	606.5	20.2	423	4	US-09-073-010-142	Sequence 142, App
37	450.5	15.0	943	4	US-09-477-135A-131	Sequence 131, App
38	424	14.1	141	4	US-09-073-009-15	Sequence 15, Appl
39	424	14.1	141	4	US-09-073-010-15	Sequence 15, Appl
40	381.5	12.7	204	4	US-08-311-731A-57	Sequence 57, Appl
41	377.5	12.6	208	4	US-08-311-731A-208	Sequence 208, App
42	328.5	10.9	371	4	US-09-050-739-92	Sequence 92, Appl
43	325.5	10.8	368	3	US-08-818-112-114	Sequence 114, App
44	325.5	10.8	368	4	US-08-818-111-109	Sequence 109, App
45	325.5	10.8	368	4	US-09-056-556-114	Sequence 114, App

ALIGNMENTS

RESULT 1
US-09-287-849-26
; Sequence 26, Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillot, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 26
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
US-09-287-849-26

Query Match		100.0%;	Score 3007;	DB 4;	Length 596;
Best Local Similarity		100.0%;	Pred. No. 1.3e-219;	Indels 0;	Gaps 0;
Matches 596;		Conservative 0;	Mismatches 0;		
Qy	1	HMEHHHHMVDFFGALPPEIN	ARMYAGPGSASISLVAAQWDSV	ASDLFSAASAFQSVVWG	60
Db	1	HMEHHHHMVDFFGALPPEIN	ARMYAGPGSASISLVAAQWDSV	ASDLFSAASAFQSVVWG	60
Qy	61	LTVGSMIGSAGLMVAASPYVAMSV	TAGQELTAAQVRVAAAAYETAYGL	TPVPVIA	120
Db	61	LTVGSMIGSAGLMVAASPYVAMSV	TAGQELTAAQVRVAAAAYETAYGL	TPVPVIA	120
Qy	121	ENRAELMILATNLGONTPALAVNEA	BCEWMAQDAALMFVAAAATATAT	ALLPFEEA	180

Db 121 ENRAELMILIAATNLGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEEA 180
Qy 181 PENTSAGGLLEQAAAEEASDTAAANQOLMNNVPQALQOLAQTOGTTPSSKLGGLWKTIVS 240
Db 181 PENTSAGGLLEQAAAEEASDTAAANQOLMNNVPQALQOLAQTOGTTPSSKLGGLWKTIVS 240
Qy 241 PHSPTSNMYSMANNHMTNSGVSWTNTLSMLKGFAPAAAQAQVOTAAQNGVRAMSSL 300
Db 241 PHSPTSNMYSMANNHMTNSGVSWTNTLSMLKGFAPAAAQAQVOTAAQNGVRAMSSL 300
Qy 301 GSSLGSSGLGGVAANLGRAASVGSLSVPQAWAAANQAVTPPAARALPLTSLTSAAEERGPG 360
Db 301 GSSLGSSGLGGVAANLGRAASVGSLSVPQAWAAANQAVTPPAARALPLTSLTSAAEERGPG 360
Qy 361 QMLGGLPVGQMGARAGGLSGVLRVPPRPVYMPHSPAAAGDIAPPALSDQRFADFPALPLD 420
Db 361 QMLGGLPVGQMGARAGGLSGVLRVPPRPVYMPHSPAAAGDIAPPALSDQRFADFPALPLD 420
Qy 421 PSAMVAQVGVQVNNINTKLYNNNAVAGGTGIVIDPENGVLVLTNNHVIAGATDINAFSVSGS 480
Db 421 PSAMVAQVGVQVNNINTKLYNNNAVAGGTGIVIDPENGVLVLTNNHVIAGATDINAFSVSGS 480
Qy 481 QTVGVVDVDRDQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMAMNSGGCGGTPRAVPG 540
Db 481 QTVGVVDVDRDQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMAMNSGGCGGTPRAVPG 540
Qy 541 RVVALGQTVQASDLSLGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
Db 541 RVVALGQTVQASDLSLGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596

RESULT 2

US-09-223-040-2
; Sequence 2, Application US/09223040
; Patent No. 6543522
; GENERAL INFORMATION:
; APPLICANT: Skeiky, Yasir
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009010US
; CURRENT APPLICATION NUMBER: US/09/223,040
; CURRENT FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-223-040-2

Query Match 97.8%; Score 2941; DB 4; Length 729;
Best Local Similarity 99.8%; Pred. No. 1.7e-214; Indels 0; Gaps 0;
Matches 587; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 9 MVDFGALPPEINSARMYAGPGSASLVAAQMDVSDVADLFSAAAFQSVVWGLTVGSWIG 68
Db 142 MVDFGALPPEINSARMYAGPGSASLVAAQMDVSDVADLFSAAAFQSVVWGLTVGSWIG 201
Qy 69 SSAGLMVAASPPVYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 128
Db 202 SSAGLMVAASPPVYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 261
Qy 129 LIATNLIGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGG 188
Db 262 LIATNLIGQNTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEEAPEMTSAGG 321
Qy 189 LLEQAAAEEASDTAAANQOLMNNVPQALQOLAQTOGTTPSSKLGGLWKTIVSPHSPISN 248

Db 322 LLEQAAAEEASDTAAANQOLMNNVPQALQOLAQTOGTTPSSKLGGLWKTIVSPHSPISN 381
Qy 249 MYSVANNHMTNSGVSWTNTLSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSG 308
Db 382 MYSVANNHMTNSGVSWTNTLSMLKGFAPAAAQAQVOTAAQNGVRAMSSLSGSSG 441
Qy 309 LGGVAAANLGRAASVGSLSVPQAWAAANQAVTPPAARALPLTSLTSAAEERGPGQMLGGLPV 368
Db 442 LGGVAAANLGRAASVGSLSVPQAWAAANQAVTPPAARALPLTSLTSAAEERGPGQMLGGLPV 501
Qy 369 GQMGARAGGLSGVLRVPPRPVYMPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 428
Db 502 GQMGARAGGLSGVLRVPPRPVYMPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 561
Qy 429 GPQVNNINTKLYNNNAVAGGTGIVIDPENGVLVLTNNHVIAGATDINAFSVSGSGQTYGVDDV 488
Db 562 GPQVNNINTKLYNNNAVAGGTGIVIDPENGVLVLTNNHVIAGATDINAFSVSGSGQTYGVDDV 621
Qy 489 GYDRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMAMNSGGCGGTPRAVPGRVWALGQT 548
Db 622 GYDRTQDVAVLQLRGAGGLPSAAICGGVAVGEPVVMAMNSGGCGGTPRAVPGRVWALGQT 681
Qy 549 VOASDLSLGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 596
Db 682 VOASDLSLGAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVGMMNTAAS 729

RESULT 3

US-09-287-849-2
; Sequence 2, Application US/09287849
; Patent No. 6827198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 729
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
US-09-287-849-2

Query Match 97.8%; Score 2941; DB 4; Length 729;
Best Local Similarity 99.8%; Pred. No. 1.7e-214; Indels 0; Gaps 0;
Matches 587; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 9 MVDFGALPPEINSARMYAGPGSASLVAAQMDVSDVADLFSAAAFQSVVWGLTVGSWIG 68
Db 142 MVDFGALPPEINSARMYAGPGSASLVAAQMDVSDVADLFSAAAFQSVVWGLTVGSWIG 201
Qy 69 SSAGLMVAASPPVYVWMSVTAGQAEITAAQVRVAAAAYETAYGLTVPPIAENRAELMI 128

Db 202 SSAGLWVAASPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 261
Qy 129 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAEPMTSAGG 188
Db 262 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAEPMTSAGG 321
Qy 189 LLEQAAAVEASDTAAANQLMNNVPQALQOQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 248
Db 322 LLEQAAAVEASDTAAANQLMNNVPQALQOQLAQPTQGTTPSSKLGGLWKTVPSPHSPISN 361
Qy 249 MVSNNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQAQNGVRAAMSSLSGSSG 308
Db 382 MVSNNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQAQNGVRAAMSSLSGSSG 441
Qy 309 LGGVAANLGRAASVGSLSVPOQAAANQAVTTPAARALPLTSLTSAERPGQMLGLPV 368
Db 442 LGGVAANLGRAASVGSLSVPOQAAANQAVTTPAARALPLTSLTSAERPGQMLGLPV 501
Qy 369 GOMCARAGGLSGVLRVPPRPYVPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 428
Db 502 GOMCARAGGLSGVLRVPPRPYVPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 561
Qy 429 GPOVNTNKLGYNNVAGAGTGVIDPENGVLNNHVIAGATDINAFSVSGSGGTGYGVVV 488
Db 562 GPOVNTNKLGYNNVAGAGTGVIDPENGVLNNHVIAGATDINAFSVSGSGGTGYGVVV 621
Qy 489 GYDRTQDAVQLRGAGLPSAIGGVAVGEPVVMGNSGGGGTTPRAVGRVVALGOT 548
Db 622 GYDRTQDAVQLRGAGLPSAIGGVAVGEPVVMGNSGGGGTTPRAVGRVVALGOT 681
Qy 549 VQASDLSLTGAETLNGLIQDAIQDGSQGVVNGLGQVVGMMNTAAS 596
Db 682 VQASDLSLTGAETLNGLIQDAIQDGSQGVVNGLGQVVGMMNTAAS 729

RESULT 4

US-09-287-849-22
; Sequence 22 Application US/09287849
; Patent No. 6627198
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: 014058-009020US
; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 22
; LENGTH: 600
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:tri-fusion
US-09-287-849-22

Query Match 57.6%; Score 2033.5; DB 4; Length 600;
Best Local Similarity 72.5%; Pred. No. 6.7e-146;

Matches 440; Conservative 19; Mismatches 81; Indels 67; Gaps 10;
Qy 1 HNHHEHHHMDGALPPPEINSARMYAGPSSASLVAAAQWDSVASDLPSAASAFQSVVWG 60
Db 1 HNHHEHHHMDGALPPPEINSARMYAGPSSASLVAAAQWDSVASDLPSAASAFQSVVWG 60
Qy 61 LTVGSWIGSSAGLVMAAASPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPVPPVIA 120
Db 61 LTVGSWIGSSAGLVMAAASPYVAMSVTAGAELTAAQVRVAAAAYETAYGLTVPVPPVIA 120
Qy 121 ENRAELMLIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEA 180
Db 121 ENRAELMLIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEA 180
Qy 181 PEMSAGLLEQAAAVEASDTAAANQLMNNVPQALQOQLAQPTQGTTPSSKLGGLWKTVS 240
Db 181 PEMSAGLLEQAAAVEASDTAAANQLMNNVPQALQOQLAQPTQGTTPSSKLGGLWKTVS 240
Qy 241 PHRSPI SNMVSNNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQAQNGVRAAMSSL 300
Db 241 PHRSPI SNMVSNNHMTNSGVSMTNTLSMLKGFAPAAAQAVQTAQAQNGVRAAMSSL 300
Qy 301 GSSLGSSGLGGVVAANLGRAASVGSLSVPOQAAANQAVTTPAARALPLTSLTSAERPG 360
Db 301 GSSLGSSGLGGVVAANLGRAASVGSLSVPOQAAANQAVTTPAARALPLTSLTSAERPG 360
Qy 361 QMLGLPLVQMGARAGGLSGVLRVPPRPYVPHSPAAAGDIAP 403
Db 361 QMLGLPLVQMGARAGGLSGVLRVPPRPYVPHSPAAAGDIAP 420
Qy 404 -PALSDQRFADFPALPLDPS 448
Db 421 LNATDPGAAQFNAPSVQSVLRNFLAAPPORAAAQQL-QAVPGAQYIGLVESVAGS 479
Qy 449 TGVIDPENGVLNNHVIAGATDINAFSVSGSGGTGYGVVVNGLGQVVGMMNTAAS 508
Db 480 CN-----NYELMTINYQFG---DVDA---HGMTRAAQASLEAEHQAIVRDLAAGDFW 527
Qy 509 SAAICGGVAVGEPVVMG-----NSGGGGTTPRAVGRVVALGOTVQASDLSLTGA 558
Db 528 GGA--GSVACQEFITQLGRNFQVIYEQANAHGQ-----KVOAGNNNAQTDSAVGS 576
Qy 559 E-ETLNG 564
Db 577 SWATNG 583

RESULT 5

US-08-818-112-107
; Sequence 107 Application US/08818112
; Patent No. 6290969
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedvick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 153
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; City: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818.112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-112-107

Query Match 64.8%; Score 1949; DB 3; Length 391;
Best Local Similarity 100.0%; Pred. No. 9.4e-140;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 MVDGALPPEINSAARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
DB 1 MVDGALPPEINSAARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 69 SSAGLWVAASPYVWMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
DB 61 SSAGLWVAASPYVWMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 129 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLPPFEAPEMTSAGG 188
DB 121 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATATLPPFEAPEMTSAGG 180
QY 189 LLEQAAVEASDSTAANQLMNNVPOALQQAQPTGGTTPSSKLGGLWKTVPSPHSPISN 248
DB 181 LLEQAAVEASDSTAANQLMNNVPOALQQAQPTGGTTPSSKLGGLWKTVPSPHSPISN 240
QY 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSLGSSG 308
DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSLGSSG 300
QY 309 LGGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAABERGQMLGGLPV 368
DB 301 LGGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAABERGQMLGGLPV 360
QY 369 GOMGARAGGGLSGVLRVPPRYVWPHSPAAG 399
DB 361 GOMGARAGGGLSGVLRVPPRYVWPHSPAAG 391

RESULT 6
US-08-818-111-102
Sequence 102, Application US/08818111
Patent No. 6338852
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonia
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 148
CORRESPONDENCE ADDRESS:
ADDRESS: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: USA

ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818.111
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.417C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 102:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-818-111-102

Query Match 64.8%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 9.4e-140;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 MVDGALPPEINSAARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
DB 1 MVDGALPPEINSAARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60
QY 69 SSAGLWVAASPYVWMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
DB 61 SSAGLWVAASPYVWMSVTAQAEELTAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120
QY 129 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLPPFEAPEMTSAGG 188
DB 121 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGVAAATATATATLPPFEAPEMTSAGG 180
QY 189 LLEQAAVEASDSTAANQLMNNVPOALQQAQPTGGTTPSSKLGGLWKTVPSPHSPISN 248
DB 181 LLEQAAVEASDSTAANQLMNNVPOALQQAQPTGGTTPSSKLGGLWKTVPSPHSPISN 240
QY 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSLGSSG 308
DB 241 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAAQAVQTAQAQNGVRAMSSLGSLGSSG 300
QY 309 LGGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAABERGQMLGGLPV 368
DB 301 LGGGVAANLGRAASVGSLSVPOQAAANQAVTPAARALPLTSLTSAABERGQMLGGLPV 360
QY 369 GOMGARAGGGLSGVLRVPPRYVWPHSPAAG 399
DB 361 GOMGARAGGGLSGVLRVPPRYVWPHSPAAG 391

RESULT 7
US-09-056-556-107
Sequence 107, Application US/09056556
Patent No. 6350456
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THE PREVENTION AND
NUMBER OF SEQUENCES: 241
CORRESPONDENCE ADDRESS:
ADDRESS: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington

```

; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/056,556
; FILING DATE: 07-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.457
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 107:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-056-556-107

```

```

Query Match 64.8%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 9.4e-140; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 0;

QY 9 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
DB 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60

QY 69 SSAGLMVAASPVVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
DB 61 SSAGLMVAASPVVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120

QY 129 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 188
DB 121 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180

QY 189 LLEQAAAVEASDTAAANQMNVPQALQQAQPTGTTTSSKLGGLWKTVPSPHSPISN 248
DB 181 LLEQAAAVEASDTAAANQMNVPQALQQAQPTGTTTSSKLGGLWKTVPSPHSPISN 240

QY 249 MYSMANNHSMNTSGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSSLSGSSG 308
DB 241 MYSMANNHSMNTSGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSSLSGSSG 300

QY 309 LGGGVAANLGRAASVGSLSVPAQAAAQAVTPAARALPLTSLTSAABERGPGQMLGLPV 368
DB 301 LGGGVAANLGRAASVGSLSVPAQAAAQAVTPAARALPLTSLTSAABERGPGQMLGLPV 360

QY 369 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 399
DB 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391

```

```

RESULT 8
US-09-072-596-102
; Sequence 102, Application US/09072596
; Patent No. 6458366
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, David C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; APPLICANT: Lodes, Michael J.
; APPLICANT: Hendrickson, Ronald C.

```

```

; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 350
; CORRESPONDENCE ADDRESS:
; ADDRESS: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/072,596
; FILING DATE: 05-MAY-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C9
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 102:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 391 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-09-072-596-102

```

```

Query Match 64.8%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 9.4e-140; Indels 0; Gaps 0;
Matches 391; Conservative 0; Mismatches 0;

QY 9 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
DB 1 MVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 60

QY 69 SSAGLMVAASPVVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPVIAENRAELMI 128
DB 61 SSAGLMVAASPVVAMSVTAGQAELETAQVRVAAAAYETAYGLTVPPVIAENRAELMI 120

QY 129 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 188
DB 121 LIATNLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLPPPEAPMTSAGG 180

QY 189 LLEQAAAVEASDTAAANQMNVPQALQQAQPTGTTTSSKLGGLWKTVPSPHSPISN 248
DB 181 LLEQAAAVEASDTAAANQMNVPQALQQAQPTGTTTSSKLGGLWKTVPSPHSPISN 240

QY 249 MYSMANNHSMNTSGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSSLSGSSG 308
DB 241 MYSMANNHSMNTSGVSMNTLSSMLKGFAPAAAQAVTAAQNGVRAMSSLSGSSG 300

QY 309 LGGGVAANLGRAASVGSLSVPAQAAAQAVTPAARALPLTSLTSAABERGPGQMLGLPV 368
DB 301 LGGGVAANLGRAASVGSLSVPAQAAAQAVTPAARALPLTSLTSAABERGPGQMLGLPV 360

QY 369 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 399
DB 361 GQMGARAGGGLSGVLRVPPRPYVMPHSPAAG 391

```

```

RESULT 9
US-09-072-967-107
; Sequence 107, Application US/09072967
; Patent No. 6592877
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.

```

APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
APPLICANT: Ledes, Michael J.
APPLICANT: Hendrickson, Ronald C.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 355
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/072.967
FILING DATE: 05-MAY-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 107:
SEQUENCE CHARACTERISTICS:
LENGTH: 391 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-072-967-107

Query Match 64.8%; Score 1949; DB 4; Length 391;
Best Local Similarity 100.0%; Pred. No. 9.4e-140;
Matches 391; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 MVDGALPPEINSAARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 68
DB 1 MVDGALPPEINSAARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 60
QY 69 SSAGLMVAAASPYVWMSVTAGQELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 128
DB 61 SSAGLMVAAASPYVWMSVTAGQELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 120
QY 129 LIATNLLGONTPAIVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 188
DB 121 LIATNLLGONTPAIVNEAEYGEWMAQDAAMFGYAAATATATATATLLPFEAPEMTSAGG 180
QY 189 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTGTTSSKLGGLWKTVPVHRSPISN 248
DB 181 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTGTTSSKLGGLWKTVPVHRSPISN 240
QY 249 MVMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSS 308
DB 241 MVMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSS 300
QY 309 LGGGVAANLGRAASVGSLSVFPQAAANQAVTPAARALPLTSLTSAARPGQMLGLPV 368
DB 301 LGGGVAANLGRAASVGSLSVFPQAAANQAVTPAARALPLTSLTSAARPGQMLGLPV 360
QY 369 GQMGARAGGGLSGVLRVPPRYPVMPHSPAAG 399
DB 361 GQMGARAGGGLSGVLRVPPRYPVMPHSPAAG 391

RESULT 10
US-08-818-112-111
Sequence 111, Application US/08818112
Patent No. 6290969
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Skeiky, Yasir A.W.
APPLICANT: Dillon, Davin C.
APPLICANT: Campos-Neto, Antonio
APPLICANT: Houghton, Raymond
APPLICANT: Vedvick, Thomas S.
APPLICANT: Twardzik, Daniel R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
AND DIAGNOSIS OF TUBERCULOSIS
NUMBER OF SEQUENCES: 153
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
CITY: Seattle
STATE: Washington
COUNTRY: USA
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/818.112
FILING DATE: 13-MAR-1997
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.411C6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 111:
SEQUENCE CHARACTERISTICS:
LENGTH: 396 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-818-112-111

Query Match 55.0%; Score 1652.5; DB 3; Length 396;
Best Local Similarity 84.9%; Pred. No. 2.7e-117;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
QY 9 MVDGALPPEINSAARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 68
DB 1 MVDGALPPEINSAARMYAGPGSASLVAAQWDSVADLFSASAFQSVVWGLTVGSGWIG 60
QY 69 SSAGLMVAAASPYVWMSVTAGQELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 128
DB 61 SSAGLMVAAASPYVWMSVTAGQELTAAQVRVAAAAYETAYGLTVPVPPVIAENRAELMI 120
QY 129 LIATNLLGONTPAIVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEAPEMTSAGG 188
DB 121 LIATNLLGONTPAIVNEAEYGEWMAQDAAMFGYAAATATATATALLPFEAPITNPGG 180
QY 189 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTGTTSSKLGGLWKTVPVHRSPISN 248
DB 181 LLEQAAVEEADTAAANQLMNNVPQALQQAQPTKSIWPFQDLSLWKALSPHLSPLSN 240
QY 249 MVMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSS 304
DB 241 IVSMLNHNHMTNSGVSMNTLHSMKGFAP-AAQAQAVETAQNGVQAMSSLSGSLGSSL 299
QY 305 GSSGLGGGVAANLGRAASVGSLSVFPQAAANQAVTPAARALPLTSLTSAARPGQMLG 364


```

Db      300 GSSGLGAGVAAANLGRRAASVGSLSVPQAWAANAQAVTPPAARALPLTSLTSAAQTAPGHMLG 399
Qy      365 GLFVGQMGARAG--GGSLGVLRVPPRPYPMPHSPAAG 399
          |||:|||:
Db      360 GLFLGQLTNSGGFGGVSNALRMPPRAYMVRVPAAG 396
          |||:|||:

RESULT 11
US-08-818-111-106
; Sequence 106, Application US/08818111
; Patent No. 6338852
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Campos-Neto, Antonia
; APPLICANT: Houghton, Raymond
; APPLICANT: Vedwick, Thomas S.
; APPLICANT: Twardzik, Daniel R.
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF TUBERCULOSIS
; NUMBER OF SEQUENCES: 148
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/818,111
; FILING DATE: 13-MAR-1997
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.417C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 106:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 396 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
US-08-818-111-106

Query Match 55.0%; Score 1652.5; DB 4; Length 396;
Best Local Similarity 84.9%; Pred. No. 2.7e-117;
Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3

Qy      9 MVDFGALPPEINSARMYAGPGSASLVAAAQWMDVASDLFSAASFQSVVMGLTVGSGWG 68
Db      1 VVDFGALPPEINSARMYAGPGSASLVAAAQWMDVASDLFSAASFQSVVMGLTVGSGWG 60
Qy      69 SSAGLMVAASAAPYVAWMSTVTAQELTAAQVRAAAAYETAYGLTYPPPVIENRAELMI 128
Db      61 SSAGLMVAASAAPYVAWMSTVTAQELTAAQVRAAAAYETAYGLTYPPPVIENRAELMI 120
Qy      129 LIATNLLGNTPTAIANEAEYGEWWAQDAAMAFGYAAATATATATLLPTEEAPEMTSAGG 188
Db      121 LIATNLLGNTPTAIANEAEYGEWWAQDAAMAFGYAAATAATAATEALLPFEDAPLLINPGG 180
Qy      189 LLEQAANVEASTAAANQLMNPNYPAOLQAQOTQCTTSPSSKLGWLKWTVPSPHRPSIN 248
Db      181 LLEQAQAVEEADITAAANQLMNPNYPAOLQAQOTKSIWPFDQISELWKAISEPHLSPLSN 240
Qy      249 MYSMANNHWSMTNSGVSMNTLTSSMLKGFAFAAAAAOAVTAAOONGVRASS----LGSSL 304

```

QY 249 MVSNNHMTNMGVMTNTLSSMLKGFAPAAAQAQVOTAAQNGVAMSS-----LGSSSL 304
 Db 241 IVSMLNHVMTNMGVMTNLTSLMLKGFAP-AAAQAVETAAQNGVAMSSLSQSLGSSSL 299
 QY 305 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERPGOMLG 364
 Db 300 GSSGLGAGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERPGHMLG 359
 QY 365 GLPVQMGARAG--CGLSGLVLRVPPRYVMPHSPAG 399
 Db 360 GLPLGQLTNSGGGFGVSNALRMPPRAYVMPRVPAAG 396
 RESULT 13
 US-09-072-596-106
 ; Sequence 106, Application US/09072596
 ; Patent No. 6458366
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonia
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Hendrickson, Ronald C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS OF
 ; TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 350
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED AND BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072,596
 ; FILING DATE: 05-MAY-1998
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.417C9
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 106:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 396 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; US-09-072-596-106
 Query Match 55.0%; Score 1652.5; DB 4; Length 396;
 Best Local Similarity 84.9%; Pred. No. 2.7e-117;
 Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
 QY 9 MVDGALPPEINARMYAGPGSASLVAAQAQWDSVADLSFAASAFQSVVWGLTVGSMIG 68
 Db 1 VVDFGALPPEINARMYAGPGSASLVAAQAQWDSVADLSFAASAFQSVVWGLTVGSMIG 60
 QY 69 SSAGLWVAASAPFYVAMSVTAGQAEELTAQVRAAAAYETAYGLTVPVPIAENRAELMI 128
 Db 61 SSAGLWVAASAPFYVAMSVTAGQAEELTAQVRAAAAYETAYGLTVPVPIAENRAELMI 120
 QY 129 LIATNLLGNTPALIAVNEAEYGEWMAQDAAMAFGAAATATATATALLPFEEAPENTSAGG 188

Db 121 LIATNLLGNTPALIAVNEAEYGEWMAQDAAMAFGAAATATATALLPFEEAPLITNPGG 180
 QY 189 LLEQAAVVEEASDTAAANQLMNNVPOALQOLAQPTQGTTPSSKLGGLWKTVPSPHRSISN 248
 Db 181 LLEQAAVVEEASDTAAANQLMNNVPOALQOLAQPTQGTTPSSKLGGLWKTVPSPHRSISN 240
 QY 249 MVSNNHMTNMGVMTNLTSLMLKGFAPAAAQAQVOTAAQNGVAMSS-----LGSSSL 304
 Db 241 IVSMLNHVMTNMGVMTNLTSLMLKGFAP-AAAQAVETAAQNGVAMSSLSQSLGSSSL 299
 QY 305 GSSGLGGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERPGOMLG 364
 Db 300 GSSGLGAGVAAANLGRAASVGSLSVPOAWAANAQVTPAARALPLTSLTSAERPGHMLG 359
 QY 365 GLPVQMGARAG--CGLSGLVLRVPPRYVMPHSPAG 399
 Db 360 GLPLGQLTNSGGGFGVSNALRMPPRAYVMPRVPAAG 396
 RESULT 14
 US-09-072-967-111
 ; Sequence 111, Application US/09072967
 ; Patent No. 6592877
 ; GENERAL INFORMATION:
 ; APPLICANT: Reed, Steven G.
 ; APPLICANT: Skeiky, Yasir A.W.
 ; APPLICANT: Dillon, Davin C.
 ; APPLICANT: Campos-Neto, Antonio
 ; APPLICANT: Houghton, Raymond
 ; APPLICANT: Vedvick, Thomas S.
 ; APPLICANT: Twardzik, Daniel R.
 ; APPLICANT: Lodes, Michael J.
 ; APPLICANT: Hendrickson, Ronald C.
 ; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY
 ; TUBERCULOSIS
 ; NUMBER OF SEQUENCES: 355
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: SEED AND BERRY LLP
 ; STREET: 6300 Columbia Center, 701 Fifth Avenue
 ; CITY: Seattle
 ; STATE: Washington
 ; COUNTRY: USA
 ; ZIP: 98104-7092
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/072,967
 ; FILING DATE: 05-MAY-1998
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Maki, David J.
 ; REGISTRATION NUMBER: 31,392
 ; REFERENCE/DOCKET NUMBER: 210121.411C9
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (206) 622-4900
 ; TELEFAX: (206) 682-6031
 ; INFORMATION FOR SEQ ID NO: 111:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 396 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: linear
 ; US-09-072-967-111
 Query Match 55.0%; Score 1652.5; DB 4; Length 396;
 Best Local Similarity 84.9%; Pred. No. 2.7e-117;
 Matches 337; Conservative 19; Mismatches 34; Indels 7; Gaps 3;
 QY 9 MVDGALPPEINARMYAGPGSASLVAAQAQWDSVADLSFAASAFQSVVWGLTVGSMIG 68

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2004, 16:39:00 ; Search time 61.2853 Seconds
(without alignments)
2747.774 Million cell updates/sec

Title: US-09-597-796C-10

Perfect score: 3007

Sequence: 1 RHHHHHHHVMDFGALPPEIN.....SGGPVNLGLGVQVGMNTAAS 596

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: A_Geneseq_29Jan04:*

1: Geneseq1980s:*

2: Geneseq1990s:*

3: Geneseq2000s:*

4: Geneseq2001s:*

5: Geneseq2002s:*

6: Geneseq2003as:*

7: Geneseq2003bs:*

8: Geneseq2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3007	100.0	596	2 AAY32070	Aay32070 Mycobacte
2	3007	100.0	596	5 AAE29710	Aae29710 Mycobacte
3	3007	100.0	596	5 AAE17574	Aae17574 Mycobacte
4	3007	100.0	599	5 AAU74599	AAU74599 Antigenic
5	2946	98.0	729	4 AAO22142	AAO22142 Ral12-H9-3
6	2946	98.0	813	7 ADA26367	ADA26367 Mycobacte
7	2946	98.0	825	7 ADA26366	ADA26366 Mycobacte
8	2946	98.0	875	7 ADA26365	ADA26365 Mycobacte
9	2946	98.0	930	5 AAE29731	AAE29731 Mycobacte
10	2946	98.0	930	7 ADA26364	ADA26364 Mycobacte
11	2946	98.0	1016	7 ADA26370	ADA26370 M. bovis
12	2946	98.0	1022	7 ADA26369	ADA26369 Mycobacte
13	2946	98.0	1154	7 ADA26368	ADA26368 Mycobacte
14	2943	97.9	729	5 AAE29709	AAE29709 Mycobacte
15	2943	97.9	729	5 AAE17573	Aae17573 Mycobacte
16	2943	97.9	729	7 ADA26374	ADA26374 Mycobacte
17	2941	97.8	729	5 AAE29708	AAE29708 Mycobacte
18	2941	97.8	729	5 AAE17572	Aae17572 Mycobacte
19	2936	97.6	729	7 ADA26373	ADA26373 Mycobacte
20	2923	97.2	729	2 AAY32059	Aay32059 Mycobacte
21	2894.5	96.3	726	5 AAU74588	AAU74588 Antigenic
22	2033.5	67.6	600	2 AAY32068	Aay32068 Mycobacte
23	2033.5	67.6	600	5 AAU74597	AAU74597 Antigenic
24	1987.5	66.1	788	4 AAU01903	AAU01903 M. tuberc
25	1960	65.2	744	4 AAU01902	AAU01902 M. tuberc

26	1960	65.2	815	4 AAU01904	AAU01904 M. tuberc
27	1955	65.0	1010	7 ADA26356	Ada26356 Mycobacte
28	1949	64.8	391	2 AAU32381	AAU32381 Mycobacte
29	1949	64.8	391	2 AAU32449	AAU32449 Mycobacte
30	1949	64.8	391	2 AAU64335	AAU64335 Mycobacte
31	1949	64.8	391	2 AAU81702	AAU81702 M. tuberc
32	1949	64.8	391	2 AAU04778	AAU04778 Mycobacte
33	1949	64.8	391	2 AAY38989	Aay38989 M. tuberc
34	1949	64.8	391	2 AAY39132	Aay39132 M. tuberc
35	1949	64.8	391	4 AAU01888	AAU01888 M. tuberc
36	1949	64.8	391	5 AAE29707	Aae29707 Mycobacte
37	1949	64.8	391	5 AAE17571	Aae17571 Mycobacte
38	1949	64.8	394	2 AAU04779	AAU04779 Mycobacte
39	1949	64.8	723	7 ADA26354	Ada26354 Mycobacte
40	1652.5	55.0	396	2 AAU64337	AAU64337 Mycobacte
41	1652.5	55.0	396	2 AAU81704	AAU81704 M. tuberc
42	1652.5	55.0	396	2 AAY38991	Aay38991 M. tuberc
43	1652.5	55.0	396	2 AAY39134	Aay39134 M. tuberc
44	1583	52.6	393	6 ABU36968	ABU36968 Protein e
45	1583	52.6	393	6 ABU34531	ABU34531 Protein e

ALIGNMENTS

RESULT 1

AAU32070
ID AAY32070 standard; protein; 596 AA.

XX AC AAY32070;

XX DT 17-JAN-2000 (first entry)

XX DE Mycobacterium tuberculosis antigen fusion protein Mtb59f.

XX KW Tuberculosis; antigen; fusion protein; Mtb59f; TbH9; Ra35; diagnosis;

XX KW therapy; vaccine; immunogen.

XX OS Mycobacterium tuberculosis.

XX XX Key Location/Qualifiers

FT Peptide 1..8

FT Protein /note= "Met/His tag"

FT Protein 9..140

FT Protein /note= "Ra12"

FT Protein 143..596

FT Protein /note= "TbH9"

XX WO9951748-A2.

XX PD 14-OCT-1999.

XX PF 07-APR-1999; 99WO-US007717.

XX PR 07-APR-1998; 98US-00056556.

XX PR 30-DEC-1998; 98US-00223040.

XX PA (CORI-) CORIYA CORP.

XX PI Skeiky YAW, Alderson M, Campos-Neto A;

XX XX WPI; 1999-601610/51.

XX XX N-PSDB; AAZ20205.

XX XX New fusion proteins useful for diagnosis, prevention and treatment of tuberculosis.

XX XX Claim 1; Fig 12A-B; 83pp; English.

XX XX This sequence represents a recombinant Mycobacterium tuberculosis bi-

XX XX antigen fusion protein, termed Mtb59f, composed of the antigens TbH9 and

XX XX Ra35. The fusion protein is expressed in host cells using a vector

XX XX carrying a polynucleotide (see AAZ20205) comprising the coding sequences

CC for the 2 antigens. The invention provides fusion proteins (see AAY32059-71) containing at least 2 M. tuberculosis antigens. The new fusion proteins and polynucleotides encoding them are useful as vaccines for preventing tuberculosis (claimed), for diagnosis (via in vitro assays or intradermal skin tests for detection of anti-M. tuberculosis antibodies), monitoring of disease progression, and treatment of tuberculosis. They are more effective immunogens than mixtures of the individual protein components

CC
XX
SQ

Query Match 100.0%; Score 3007; DB 2; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.7e-198;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HHHHHHHVDFGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWG 60
DB 1 HHHHHHHVDFGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWG 60
QY 61 LTVGSWIGSSAGLMVAASAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPVIA 120
DB 61 LTVGSWIGSSAGLMVAASAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPVIA 120
QY 121 ENRAELMILITNLGONTPAIAYNEAYGEMWAQDAAMFGYAAATATATATALLPFEA 180
DB 121 ENRAELMILITNLGONTPAIAYNEAYGEMWAQDAAMFGYAAATATATATALLPFEA 180
QY 181 PEMTSAGGLLEQAAAEEASDTAAANQLMNNVPAALQQLAQPTQGTTPSSKLGGLWKTVS 240
DB 181 PEMTSAGGLLEQAAAEEASDTAAANQLMNNVPAALQQLAQPTQGTTPSSKLGGLWKTVS 240
QY 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSL 300
DB 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSL 300
QY 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGGP 360
DB 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGGP 360
QY 361 QMLGGLPVQMGARAGGSLGVLVPPRYPMHSPAAAGDIAPPALSQDRFADFPALPLD 420
DB 361 QMLGGLPVQMGARAGGSLGVLVPPRYPMHSPAAAGDIAPPALSQDRFADFPALPLD 420
QY 421 PSAMVAQVGPVQVNTKLGYNNAVAGTGIVDPNGVLTNNHVIAGTDINAFSVGSG 480
DB 421 PSAMVAQVGPVQVNTKLGYNNAVAGTGIVDPNGVLTNNHVIAGTDINAFSVGSG 480
QY 481 QTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPG 540
DB 481 QTYGVDVVGVDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPG 540
QY 541 RVVAGQTVQASDSLTAETGLNGLIQPDAAIQGDSGGPVWNGLGQVGMNTAAS 596
DB 541 RVVAGQTVQASDSLTAETGLNGLIQPDAAIQGDSGGPVWNGLGQVGMNTAAS 596

RESULT 2
AAE29710
ID AAE29710 standard; protein; 596 AA.
XX
AC AAE29710;
XX
DT 29-AUG-2003 (revised)
DT 27-JAN-2003 (first entry)
XX
XX Mycobacterium sp. MTB59F fusion protein.
DE Vaccine; immunity; diagnostic agent; gene therapy; TBH9; antigen; Ra35;
KW MTB59F; fusion protein.
XX
XX Mycobacterium sp.
OS Mycobacterium tuberculosis.
OS Chimeric.

XX WO200272792-A2.
XX
PD 19-SEP-2002.
XX 13-MAR-2002; 2002WO-US008223.
XX 13-MAR-2001; 2001US-0275837P.
XX (CORI-) CORIXA CORP.
XX
XX Skeiky, Y, Brannon M, Guderian J;
XX WPI; 2002-759944/82.
DR N-PSDB; AAD47086.
XX
XX New recombinant nucleic acid molecule comprising a Leishmania TSA, Leif, M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity against pathogenic microorganisms e.g. Leishmania and Mycobacterium tuberculosis.
XX
XX Disclosure; Page 98-99; 155pp; English.
XX
XX The invention relates to a recombinant nucleic acid molecule encoding a fusion polypeptide. The recombinant nucleic acid comprises a heterologous polynucleotide sequence encoding an antigen or an antigenic fragment from Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a polypeptide or its fragment. The Leishmania polynucleotide is selected from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention are used in methods for eliciting immune response in mammals. They are useful as vaccines to elicit protective immunity against pathogenic microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion polypeptides are used for enhancing the expression of polynucleotides, as in vivo diagnostic agents and for raising antibodies in a non-human animal. The invention is used in gene therapy. The present sequence is MTB59F fusion protein. This fusion protein comprises Ra35 protein from Mycobacterium tuberculosis and TBH9 protein from Mycobacterium sp. (Updated on 29-AUG-2003 to standardise OS field)

XX
SQ

Query Match 100.0%; Score 3007; DB 5; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.7e-198;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 HHHHHHHVDFGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWG 60
DB 1 HHHHHHHVDFGALPEINSARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWG 60
QY 61 LTVGSWIGSSAGLMVAASAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPVIA 120
DB 61 LTVGSWIGSSAGLMVAASAPYVAMSVTAGQAEITAAQVRVAAAAYETAYGLTVPVPVIA 120
QY 121 ENRAELMILITNLGONTPAIAYNEAYGEMWAQDAAMFGYAAATATATATALLPFEA 180
DB 121 ENRAELMILITNLGONTPAIAYNEAYGEMWAQDAAMFGYAAATATATATALLPFEA 180
QY 181 PEMTSAGGLLEQAAAEEASDTAAANQLMNNVPAALQQLAQPTQGTTPSSKLGGLWKTVS 240
DB 181 PEMTSAGGLLEQAAAEEASDTAAANQLMNNVPAALQQLAQPTQGTTPSSKLGGLWKTVS 240
QY 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSL 300
DB 241 PHRSPISNMVSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVQTAQAQNGVRAMSSL 300
QY 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGGP 360
DB 301 GSSLGSSGLGGVAAANLGRAASVGSLSVPQAWAANQAVTPAARALPLTSLTSAERGGP 360
QY 361 QMLGGLPVQMGARAGGSLGVLVPPRYPMHSPAAAGDIAPPALSQDRFADFPALPLD 420
DB 361 QMLGGLPVQMGARAGGSLGVLVPPRYPMHSPAAAGDIAPPALSQDRFADFPALPLD 420

QY 421 PSAMVAQVGPVNNINIKLGYNNVAGAGTGIVIDPVGVLTNHVIAGATDINAFSVGSG 480
Db 421 PSAMVAQVGPVNNINIKLGYNNVAGAGTGIVIDPVGVLTNHVIAGATDINAFSVGSG 480
QY 481 QTYGVVDVGYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPG 540
Db 481 QTYGVVDVGYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPG 540
QY 541 RVVALGQTVQASDSLTAETLNGLIQFDAAIQGDSGGPVVNGLGQVVGMMNTAAS 596
Db 541 RVVALGQTVQASDSLTAETLNGLIQFDAAIQGDSGGPVVNGLGQVVGMMNTAAS 596

RESULT 3
AAE17574
ID AAE17574 standard; protein; 596 AA.
XX AAE17574;
AC
XX
DT 22-APR-2002 (first entry)
XX
DE Mycobacterium species MTB59F fusion protein.
XX
XX Fusion protein; antigen; serological sensitivity; immune response;
KW tuberculosis; infection; vaccine; MTB59F; TbH9-Ra35 protein.
XX
OS Mycobacterium sp.
XX
XX WO200198460-A2.
PN
XX
PD 27-DEC-2001.
XX
XX 20-JUN-2001; 2001WO-US019959.
XX
XX 20-JUN-2000; 2000US-00597796.
PR
PR 01-FEB-2001; 2001US-0265737P.
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Skeiky Y, Reed S, Alderson M;
PI
XX
XX WPI; 2002-147798/19.
DR
DR N-PSDB; AAD28344.
XX
XX Composition comprising MTB39 antigen and MTB32A antigen from
PT Mycobacterium species, useful for eliciting immune response in a subject.
FT
XX
XX Claim 5; Page 114-115; 136pp; English.

CC The present invention relates to fusion proteins containing at least two
CC Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected with
CC Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
CC MTB59F (TbH9-Ra35) fusion protein

XX Sequence 596 AA;

Query Match 100.0%; Score 3007; DB 5; Length 596;
Best Local Similarity 100.0%; Pred. No. 1.7e-198;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 HMEHHHHHMFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLPSAASAFQSVVWG 60
Db 1 HMEHHHHHMFVDFGALPPEINSARMYAGPGSASLVAAQWDSVASDLPSAASAFQSVVWG 60
QY 61 LTVGSWIGSSAGLMVAAASPYVAMSVTAGAEITAAQVRVAAAAYETAYGLTVPPIVIA 120
Db 61 LTVGSWIGSSAGLMVAAASPYVAMSVTAGAEITAAQVRVAAAAYETAYGLTVPPIVIA 120
QY 121 ENRAELMILIAINLLGONTIPAIVNEAEYGEWQAQDAAMEGYAAATATATATLLPPEEA 180
Db 121 ENRAELMILIAINLLGONTIPAIVNEAEYGEWQAQDAAMEGYAAATATATATLLPPEEA 180
QY 181 PEMTSAGGLLECAAAVEEASDTAAANQLMNNVPQALQLAQPTGTTTSSKLGGLWKTVS 240
Db 181 PEMTSAGGLLECAAAVEEASDTAAANQLMNNVPQALQLAQPTGTTTSSKLGGLWKTVS 240
QY 241 PHRSPISNNVSMANNHSMNTNSGVMTNTLSSMLKGFAPAAAQAVQTAAGVGRAMSSL 300
Db 241 PHRSPISNNVSMANNHSMNTNSGVMTNTLSSMLKGFAPAAAQAVQTAAGVGRAMSSL 300
QY 301 GSSLGSSGLGGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERPG 360
Db 301 GSSLGSSGLGGGVAANLGRAASVGSLSVPOAWAANQAVTPAARALPLTSLTSAERPG 360
QY 361 QMLGGLPVGMQARAGGSLGVLRVPPRPYVMPHSIPAAGDIAPPALSQDRPADPPLD 420
Db 361 QMLGGLPVGMQARAGGSLGVLRVPPRPYVMPHSIPAAGDIAPPALSQDRPADPPLD 420
QY 421 PSAMVAQVGPVNNINIKLGYNNVAGAGTGIVIDPVGVLTNHVIAGATDINAFSVGSG 480
Db 421 PSAMVAQVGPVNNINIKLGYNNVAGAGTGIVIDPVGVLTNHVIAGATDINAFSVGSG 480
QY 481 QTYGVVDVGYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPG 540
Db 481 QTYGVVDVGYDRTQDVAVLQRLGAGGLPSAAIGGGVAVGEPVVMGNSGGQGTTPRAVPG 540
QY 541 RVVALGQTVQASDSLTAETLNGLIQFDAAIQGDSGGPVVNGLGQVVGMMNTAAS 596
Db 541 RVVALGQTVQASDSLTAETLNGLIQFDAAIQGDSGGPVVNGLGQVVGMMNTAAS 596

RESULT 4
AAU74599
ID AAU74599 standard; protein; 599 AA.
XX
AC AAU74599;
XX
DT 29-AUG-2003 (revised)
DT 08-MAY-2002 (first entry)
XX
DE Antigenic fusion protein Tb59-Ra35 (MtB59f).
XX
KW Fusion protein; tuberculosis; Mycobacterium tuberculosis;
KW tuberculostatic; immunogen; vaccine; Tb59-Ra35; MtB59f.
XX
OS Mycobacterium tuberculosis.
OS Chimeric.
XX
FH Key Location/Qualifiers
FT Misc-difference 597
FT /label= OTHER
FT /note= "OTHER= Xaa. Xaa= In frame stop codon"
XX
PN US2002009459-A1.
XX
XX 24-JAN-2002.
XX
XX 07-APR-1999; 99US-00287849.
XX

13-MAR-1997; 97US-00818112.
01-OCT-1997; 97US-00942578.
18-FEB-1998; 98US-00025197.
07-APR-1998; 98US-00056556.
30-DEC-1998; 98US-00223040.
(REED/) REED S G.
(SKEI/) SKEIKY Y A.
(DILL/) DILLON D C.
(ALDE/) ALDERSON M.
(CAMP/) CAMPOS-NETO A.
Reed SG, Skeiky YA, Dillon DC, Alderson M, Campos-Neto A;
WPI; 2002-171134/22.
N-PSDB; ABK14139.
New fusion proteins of Mycobacterium tuberculosis antigens, useful for
diagnosing, treating or preventing M. tuberculosis infection,
particularly as vaccine for treating or preventing tuberculosis.
Claim 1; Fig 12; 62pp; English.
The invention relates to a purified polypeptide which induces an immune
response of Mycobacterium tuberculosis. Polypeptides of the invention are
useful for diagnosing, treating or preventing M. tuberculosis infection,
particularly tuberculosis infection. In particular, the polypeptides are
useful as a vaccine formulation with an adjuvant to afford long-term
protection in animals against the development of tuberculosis. The
protein coding sequence may be used to encode a protein product for use
as an immunogen to induce and/or enhance an immune response to M.
tuberculosis. This sequence represents an M. tuberculosis fusion protein
of the invention. (Updated on 29-AUG-2003 to standardise OS field)
SQ Sequence 599 AA;
Query Match 100.0%; Score 3007; DB 5; Length 599;
Best Local Similarity 100.0%; Pred. No. 1.7e-198;
Matches 596; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 HMEHHHMHVDFGALPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWG 60
DB 1 HMEHHHMHVDFGALPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWG 60
QY 61 LTGSGWIGSSAGLVAAASPTVAMSVTAGAELTAAQVRVAAAYETAYGLTVPPPVIA 120
DB 61 LTGSGWIGSSAGLVAAASPTVAMSVTAGAELTAAQVRVAAAYETAYGLTVPPPVIA 120
QY 121 ENRAELMILITNLGQNTPAIYNEAYGEMWAOADAAMFGVAAATATATATILLPEEA 180
DB 121 ENRAELMILITNLGQNTPAIYNEAYGEMWAOADAAMFGVAAATATATATILLPEEA 180
QY 181 PEMSAGLLQAAVEASDAAANQIMNNVPAALQQAQPTQGTQTPSSKLGGLWKTVS 240
DB 181 PEMSAGLLQAAVEASDAAANQIMNNVPAALQQAQPTQGTQTPSSKLGGLWKTVS 240
QY 241 PHSPISNMVMANNHMTSGVSMYTLSSMLKGFAPAAAQAVTAQNGVRAMSSL 300
DB 241 PHSPISNMVMANNHMTSGVSMYTLSSMLKGFAPAAAQAVTAQNGVRAMSSL 300
QY 301 GSSLGSSGLGGVAANLGRAASVGSLSVPOAAANQAVTPAARALPLTSLTSAERGGP 360
DB 301 GSSLGSSGLGGVAANLGRAASVGSLSVPOAAANQAVTPAARALPLTSLTSAERGGP 360
QY 361 QMLGSLPVQMGARAGGSLGVLVPRPPYVPHSPDAAGDIAPALSQDRPAPPLPLD 420
DB 361 QMLGSLPVQMGARAGGSLGVLVPRPPYVPHSPDAAGDIAPALSQDRPAPPLPLD 420
QY 421 PSAMVAQVGPQVNTKLGYNNAVAGCTGTVIPDNGVLTNNHVIAGATDINAFVSGS 480
DB 421 PSAMVAQVGPQVNTKLGYNNAVAGCTGTVIPDNGVLTNNHVIAGATDINAFVSGS 480
QY 481 QTYGVVDVYDRTQDVAVLQLRAGGLPSAAIGGGVAVGEPVAVMGNSGGGGTTPRAVPG 540

DB 481 QTYGVVDVYDRTQDVAVLQLRAGGLPSAAIGGGVAVGEPVAVMGNSGGGGTTPRAVPG 540
QY 541 RVVALGQTVQASDSLTGAETINGLIQFDALIQPDGSGGPPVNGLGQVVGNTAAS 596
DB 541 RVVALGQTVQASDSLTGAETINGLIQFDALIQPDGSGGPPVNGLGQVVGNTAAS 596
RESULT 5
AAO22142
ID AAO22142 standard; protein; 729 AA.
XX AC AAO22142;
XX DT 03-OCT-2002 (first entry)
XX DE Ra12-H9-32A fusion protein.
XX KW Ra12; serine protease antigen; MTB32A; Mycobacterium tuberculosis;
KW vaccine; eukaryotic; prokaryotic origin; serum antibody; fusion protein;
KW immunogen; cytokine.
XX OS Unidentified.
XX OS Chimeric.
XX PN WO200125401-A2.
XX PD 12-APR-2001.
XX PF 06-OCT-2000; 2000WO-US027652.
XX PR 07-OCT-1999; 99US-0158585P.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky Y, Guderian J;
XX WPI; 2001-266299/27.
XX N-PSDB; AAL40773.
PT Recombinant nucleic acid molecule for producing high yield expression of
desired fusion polypeptides, encodes fusion polypeptide comprising
Mycobacterium tuberculosis coding sequence and heterologous polypeptide.
XX Disclosure; Fig 6; 39pp; English.
XX The invention relates to a recombinant nucleic acid molecule encoding a
fusion polypeptide, comprising a polynucleotide sequence of Ra12, a 14
kDa C-terminal fragment of serine protease antigen MTB32A of
Mycobacterium tuberculosis, and a heterologous polynucleotide sequence.
XX The recombinant fusion nucleic acids and polypeptides are useful for
providing stable and high yield expression of fusion polypeptides of both
eukaryotic and prokaryotic origin and to encode a protein product for use
as an antigen for detecting serum antibodies. The presence of serum
antibodies to M. tuberculosis antigens in an individual indicates that
the individual is infected with it. The fusion polypeptides are useful as
sources of proteins for monitoring binding of serum antibodies to fusion
proteins and as an immunogen to induce and/or enhance immune responses.
XX The coding sequences can be ligated with a coding sequence of another
molecule such as M. tuberculosis antigens, a cytokine or an adjuvant and
can be used in vivo as a DNA vaccine. This sequence represents the Ra12-
H9-32A fusion protein
XX Sequence 729 AA;
Query Match 98.0%; Score 2946; DB 4; Length 729;
Best Local Similarity 100.0%; Pred. No. 3.4e-194;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 MYDFGALPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
DB 142 MYDFGALPEINARMYAGPGSASLVAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201

Db 322 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 381
QY 249 MVSMAHHMSTNSGVSMTNTLSMLKGFAPAAAAQAVQTAQNGVRAMSSLGSSG 308
Db 382 MVSMAHHMSTNSGVSMTNTLSMLKGFAPAAAAQAVQTAQNGVRAMSSLGSSG 441
QY 309 LGGVAAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQGLGLPV 368
Db 442 LGGVAAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQGLGLPV 501
QY 369 GQMGARAGGSLGVLRVPPRYMPSPAAGDTAPPALSDQDRPADFPALPLDPSAMVAQV 428
Db 502 GQMGARAGGSLGVLRVPPRYMPSPAAGDTAPPALSDQDRPADFPALPLDPSAMVAQV 561
QY 429 GPQVNNITKLGYNNAVAGTGVIDPNGVLTNNHVIAGATDINAFSGSGGTGYVDVV 488
Db 562 GPQVNNITKLGYNNAVAGTGVIDPNGVLTNNHVIAGATDINAFSGSGGTGYVDVV 621
QY 489 GYDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGNSGGGTPRAVPGRVVAGQT 548
Db 622 GYDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGNSGGGTPRAVPGRVVAGQT 681
QY 549 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMNTAAS 596
Db 682 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMNTAAS 729
RESULT 9
AAE29731
ID AAE29731 standard; protein; 930 AA.
XX
AC AAE29731;
XX
DT 29-AUG-2003 (revised)
DT 27-JAN-2003 (first entry)
XX
DE Mycobacterium sp. MTB72F-Leishmania sp. MAPS (aka r95f) fusion protein.
XX
KW Vaccine; immunity; diagnostic agent; gene therapy; MTB72F; MAPS;
KW fusion protein.
XX
OS Mycobacterium sp.
OS Leishmania sp.
OS Chimeric.
XX
FN WO200272792-A2.
XX
PD 19-SEP-2002.
XX
PF 13-MAR-2002; 2002WO-US008223.
XX
PR 13-MAR-2001; 2001US-0275837P.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Brannon M, Guderian J;
XX
DR WFI; 2002-759844/82.
DR N-PSDB; AAD47110.
XX
PT New recombinant nucleic acid molecule comprising a Leishmania TSA, Leif,
PT M15 or 6H polynucleotide, useful as vaccine to elicit protective immunity
PT against pathogenic microorganisms e.g. Leishmania and Mycobacterium
PT tuberculosis.
XX
PS Example 6; Page 129-132; 155pp; English.
XX
CC The invention relates to a recombinant nucleic acid molecule encoding a
CC fusion polypeptide. The recombinant nucleic acid comprises a heterologous
CC polynucleotide sequence encoding an antigen or an antigenic fragment from
CC Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
CC polypeptide or its fragment. The Leishmania polynucleotide is selected
CC from TSA, Leif, M15, and 6H polynucleotides. Sequences of the invention

CC are used in methods for eliciting immune response in mammals. They are
CC useful as vaccines to elicit protective immunity against pathogenic
CC microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
CC polypeptides are used for enhancing the expression of polynucleotides, as
CC in vivo diagnostic agents and for raising antibodies in a non-human
CC animal. The invention is used in gene therapy. The present sequence is
CC Mycobacterium sp. MTB72F-Leishmania sp. thiol specific antitoxin [TSA;
CC MAPS (aka r95f)] fusion protein. This sequence comprises Mycobacterium
CC species MTB72F (a 72 kDa poly-protein fusion construct comprising Ral2-
CC TbH9-Ra35) linked to the Leishmania TSA (MAPS). (Updated on 29-AUG-2003
CC to standardise OS field)
XX
SQ Sequence 930 AA;
Query Match 98.0%; Score 2946; DB 5; Length 930;
Best Local Similarity 100.0%; Pred. No. 4.6e-194;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 MYDFGALPPEINSAARMYAGPGSASLVAAAQMWDSVADSLFSAASAPQSVVMGLTVGSMIG 68
Db 142 MYDFGALPPEINSAARMYAGPGSASLVAAAQMWDSVADSLFSAASAPQSVVMGLTVGSMIG 201
QY 69 SSAGLWVAASPYVAMSVTAQAEELTAQVRAAAAVETAYGLTVPDPVIAEVRALMI 128
Db 202 SSAGLWVAASPYVAMSVTAQAEELTAQVRAAAAVETAYGLTVPDPVIAEVRALMI 261
QY 129 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGG 188
Db 262 LIATNLLGONTPAIAVNEAEYGEWMAQDAAMFGYAAATATATATLLPFEEAPEMTSAGG 321
QY 189 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 248
Db 322 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 381
QY 249 MVSMAHHMSTNSGVSMTNTLSMLKGFAPAAAAQAVQTAQNGVRAMSSLGSSG 308
Db 382 MVSMAHHMSTNSGVSMTNTLSMLKGFAPAAAAQAVQTAQNGVRAMSSLGSSG 441
QY 309 LGGVAAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQGLGLPV 368
Db 442 LGGVAAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQGLGLPV 501
QY 369 GQMGARAGGSLGVLRVPPRYMPSPAAGDTAPPALSDQDRPADFPALPLDPSAMVAQV 428
Db 502 GQMGARAGGSLGVLRVPPRYMPSPAAGDTAPPALSDQDRPADFPALPLDPSAMVAQV 561
QY 429 GPQVNNITKLGYNNAVAGTGVIDPNGVLTNNHVIAGATDINAFSGSGGTGYVDVV 488
Db 562 GPQVNNITKLGYNNAVAGTGVIDPNGVLTNNHVIAGATDINAFSGSGGTGYVDVV 621
QY 489 GYDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGNSGGGTPRAVPGRVVAGQT 548
Db 622 GYDRTQDVAVLQRCAGGLPSAAIGGGVAVGEPVVMGNSGGGTPRAVPGRVVAGQT 681
QY 549 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMNTAAS 596
Db 682 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVVMNTAAS 729
RESULT 10
ADA26364
ID ADA26364 standard; protein; 930 AA.
XX
AC ADA26364;
XX
DT 20-NOV-2003 (first entry)
XX
DE Mycobacterium MTB72F-MAPS (fusion r95f) protein.
KW fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW tuberculosis; tuberculostatic; gene therapy; vaccine.
XX
OS Chimeric.

OS Mycobacterium sp.
XX WO2003070187-A2.
XX 28-AUG-2003.
XX 18-FEB-2003; 2003WO-US004903.
XX 15-FEB-2002; 2002US-0357351P.
XX (CORI-) CORIXA CORP.
PI Skeiky Y, Guderian J, Reed S;
XX WPI; 2003-697554/66.
XX N-PSDB; ADA26357.
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT and MTB39 antigens, with or without the MTB85A antigen, from a
PT Mycobacterium species, useful for diagnosing, preventing and/or treating
PT tuberculosis infection.
XX Disclosure; Fig 12; 112pp; English.
XX The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The
CC present sequence is used in the exemplification of the invention.
XX Sequence 930 AA;
SQ
Query Match 98.0%; Score 2946; DB 7; Length 930;
Best Local Similarity 100.0%; Pred. No. 4.6e-194;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 MVDGALPEINSAARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
DB 142 MVDGALPEINSAARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201
QY 69 SSAGLMVAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMI 128
DB 202 SSAGLMVAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMI 261
QY 129 LIATNLLGONTPTAIVNEAEYGEWQAQDAAMFGYAAATATATATATLLPFEAPBMTSAGG 188
DB 262 LIATNLLGONTPTAIVNEAEYGEWQAQDAAMFGYAAATATATATATLLPFEAPBMTSAGG 321
QY 189 LLEQAAVEEASDTAAANQLMNNVPQALQQAQTPQTGTTSPSSKLGLWKTVPSPHRSPISN 248
DB 322 LLEQAAVEEASDTAAANQLMNNVPQALQQAQTPQTGTTSPSSKLGLWKTVPSPHRSPISN 381
QY 249 MYSNANNHMTNSGVSNTLSSMLKGFAPAAAAQVQTAQAQNGVRAMSSLGSSG 308
DB 382 MYSNANNHMTNSGVSNTLSSMLKGFAPAAAAQVQTAQAQNGVRAMSSLGSSG 441
QY 309 LGGGVAANLGRAASVGSLSVQAAANAQVTPAARALPLTSLTSAERGGQMLGLPV 368
DB 442 LGGGVAANLGRAASVGSLSVQAAANAQVTPAARALPLTSLTSAERGGQMLGLPV 501
QY 369 QMGARAGGSLGVLRVPPRYVPHSPAGDIAAPPALSDQRFADFPALPLDPSAMVAQV 428
DB 502 QMGARAGGSLGVLRVPPRYVPHSPAGDIAAPPALSDQRFADFPALPLDPSAMVAQV 561
QY 429 GPQVYNITKLYNNAGVAGTGIVDPNGVLTNNHVIAGATDINAFSGVSGGTGYDDV 488
DB 562 GPQVYNITKLYNNAGVAGTGIVDPNGVLTNNHVIAGATDINAFSGVSGGTGYDDV 621
QY 489 GYDRTQDVAVLQRLGAGGLPSAAIGGVVAGPEPVVAMGNSGGQGTTPRAVGRVVALGQT 548

DB 622 GYDRTQDVAVLQRLGAGGLPSAAIGGVVAGPEPVVAMGNSGGQGTTPRAVGRVVALGQT 681
QY 549 VOASDLSLTGAETLNGLIQFDAAIQFGDSGGPVVNGLGQVVGMMNTAAS 596
DB 682 VOASDLSLTGAETLNGLIQFDAAIQFGDSGGPVVNGLGQVVGMMNTAAS 729
RESULT 11
ADA26370
ID ADA26370 standard; protein; 1016 AA.
XX ADA26370;
XX 20-NOV-2003 (first entry)
XX M. bovis MTB72F and 85b complex antigen (fusion MTB103F).
XX fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW tuberculosis; tuberculostatic; gene therapy; vaccine; MTB72F; 85B.
XX Chimeric.
OS Mycobacterium bovis.
XX WO2003070187-A2.
XX 28-AUG-2003.
XX 18-FEB-2003; 2003WO-US004903.
XX 15-FEB-2002; 2002US-0357351P.
XX (CORI-) CORIXA CORP.
XX Skeiky Y, Guderian J, Reed S;
XX WPI; 2003-697554/66.
XX N-PSDB; ADA26363.
XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT and MTB39 antigens, with or without the MTB85A antigen, from a
PT Mycobacterium species, useful for diagnosing, preventing and/or treating
PT tuberculosis infection.
XX Claim 85; Fig 18; 112pp; English.
XX The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The
CC present sequence is used in the exemplification of the invention.
XX Sequence 1016 AA;
SQ
Query Match 98.0%; Score 2946; DB 7; Length 1016;
Best Local Similarity 100.0%; Pred. No. 5.1e-194;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 9 MVDGALPEINSAARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 68
DB 142 MVDGALPEINSAARMYAGPGSASLVAAAQWDSVASDLFSAASAFQSVVWGLTVGSMIG 201
QY 69 SSAGLMVAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMI 128
DB 202 SSAGLMVAASPYVAMSVTAQAEELTAQVRVAAAAYETAYGLTVPPIAENRAELMI 261
QY 129 LIATNLLGONTPTAIVNEAEYGEWQAQDAAMFGYAAATATATATLLPFEAPBMTSAGG 188
DB 262 LIATNLLGONTPTAIVNEAEYGEWQAQDAAMFGYAAATATATATLLPFEAPBMTSAGG 321
QY 189 LLEQAAVEEASDTAAANQLMNNVPQALQQAQTPQTGTTSPSSKLGLWKTVPSPHRSPISN 248

Db 322 LLEQAAVEEASDTAAANQNNVPOALQQAQPTQGTTFSSKLGKLTWVSPHRSPI SN 381
QY 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSLSGSSIGSSG 308
Db 382 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSLSGSSIGSSG 441
QY 309 LGCGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALPLTSLTSAARPGQMLGGLPV 368
Db 442 LGCGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALPLTSLTSAARPGQMLGGLPV 501
QY 369 GQMGARAGGSLGVLRVPPRYMPHSPAAAGDIAPPALSDRFPALPLDPSAMVAQV 428
Db 502 GQMGARAGGSLGVLRVPPRYMPHSPAAAGDIAPPALSDRFPALPLDPSAMVAQV 561
QY 429 GPQVNVINTKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDV 488
Db 562 GPQVNVINTKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDV 621
QY 489 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVMGNSSGGGGTPPRAVPGRVVALGQT 548
Db 622 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVMGNSSGGGGTPPRAVPGRVVALGQT 681
QY 549 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVGMNTAAS 596
Db 682 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVGMNTAAS 729

RESULT 12

ADA26369
ID ADA26369 standard; protein; 1022 AA.

XX ADA26369;
XX
XX
DT 20-NOV-2003 (first entry)
XX
DE Mycobacterium MTB72F-mTCC#1 (fusion MTB102tm2F) protein.
XX
XX fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW tuberculosis; tuberculostatic; gene therapy; vaccine.
XX
OS Chimeric.
OS Mycobacterium sp.

XX WO2003070187-A2.
XX

XX 28-AUG-2003.
XX

XX 18-FEB-2003; 2003WO-US004903.
XX

XX 15-FEB-2002; 2002US-0357351P.
XX

XX (CORI-) CORIXA CORP.
XX

XX Skeiky Y, Guderian J, Reed S;
XX

XX PI
XX
XX WPI; 2003-697554/66.
XX

XX N-PSDB; ADA26362.
XX

XX New isolated nucleic acid encoding a fusion polypeptide with the MTB32A
PT and MTB39 antigens, with or without the MTB85A antigen, from a
PT Mycobacterium species, useful for diagnosing, preventing and/or treating
PT tuberculosis infection.
XX

XX Claim 85; Fig 17; 112pp; English.
XX

XX The invention relates to a novel nucleic acid encoding a fusion
CC polypeptide, comprising a MTB32A and MTB39 antigen, or MTB32A, MTB39 and
CC MTB85A antigen from a Mycobacterium species of the tuberculosis complex.
CC A polypeptide of the invention has tuberculostatic activity. A
CC polynucleotide of the invention may have a use in gene therapy, and as a
CC vaccine. The methods and compositions of the present invention, are useful
CC for diagnosing, preventing and/or treating tuberculosis infection. The

CC present sequence is used in the exemplification of the invention.

XX
SQ Sequence 1022 AA;

Query Match 98.0%; Score 2946; DB 7; Length 1022;
Best Local Similarity 100.0%; Pred. No. 5.2e-194;
Matches 588; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 MVDFGALPPEINSARMYAGPGSASLVAAQAQWSDVSDLSAASAFOSVVMGLTVGWSIG 68
Db 142 MVDFGALPPEINSARMYAGPGSASLVAAQAQWSDVSDLSAASAFOSVVMGLTVGWSIG 201

QY 59 SSAGLWVAASPYVAMSVTAGOAELETAQVRVAAAAYETAYGLTVPPVTAENRAELMI 128
Db 202 SSAGLWVAASPYVAMSVTAGOAELETAQVRVAAAAYETAYGLTVPPVTAENRAELMI 261

QY 129 LIATNLGQNTPALAVNEAEYGEWMAQDAAMFYAAATATATATLTPFEAPEMTSAGG 188
Db 262 LIATNLGQNTPALAVNEAEYGEWMAQDAAMFYAAATATATLTPFEAPEMTSAGG 321

QY 189 LLEQAAVEEASDTAAANQNNVPOALQQAQPTQGTTFSSKLGKLTWVSPHRSPI SN 248
Db 322 LLEQAAVEEASDTAAANQNNVPOALQQAQPTQGTTFSSKLGKLTWVSPHRSPI SN 381

QY 249 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSLSGSSIGSSG 308
Db 382 MYSMANNHMTNSGVSMNTLSSMLKGFAPAAAQAVOTAAQNGVRAMSSLSGSSIGSSG 441

QY 309 LGCGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALPLTSLTSAARPGQMLGGLPV 368
Db 442 LGCGVAANLGRAASVGSLSVPQAWAANAQAVTPAARALPLTSLTSAARPGQMLGGLPV 501

QY 369 GQMGARAGGSLGVLRVPPRYMPHSPAAAGDIAPPALSDRFPALPLDPSAMVAQV 428
Db 502 GQMGARAGGSLGVLRVPPRYMPHSPAAAGDIAPPALSDRFPALPLDPSAMVAQV 561

QY 429 GPQVNVINTKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDV 488
Db 562 GPQVNVINTKLYNNVAVGAGTGIVDPNGVLTNNHVIAGATDINAFSVGSGQTYGVDDV 621

QY 489 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVMGNSSGGGGTPPRAVPGRVVALGQT 548
Db 622 GYDRTQDVAVLQLRGAGGLPSAAIGGGVAVGEPVAVMGNSSGGGGTPPRAVPGRVVALGQT 681

QY 549 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVGMNTAAS 596
Db 682 VQASDSLTAETLNGLIQFDAAIQPGDSGGPVVNGLGQVGMNTAAS 729

RESULT 13

ADA26368
ID ADA26368 standard; protein; 1154 AA.

XX ADA26368;
XX

XX 20-NOV-2003 (first entry)
XX

XX Mycobacterium MTB72F-mTCC#2 (fusion MTB114F) protein.
XX

XX fusion protein; MTB32A; MTB39; antigen; MTB32A; MTB39; MTB85A;
KW tuberculosis; tuberculostatic; gene therapy; vaccine.
XX

XX Chimeric.
XX

XX Mycobacterium sp.
XX

XX WO2003070187-A2.
XX

XX 28-AUG-2003.
XX

XX 18-FEB-2003; 2003WO-US004903.
XX

XX 15-FEB-2002; 2002US-0357351P.
XX


```
Db 322 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 381
Qy 249 MYSMANNHMSMTNSGVSMNTLSMLKGFAPAPAAAQAVQTAQNGVRAMSSLGSSG 308
Db 382 MYSMANNHMSMTNSGVSMNTLSMLKGFAPAPAAAQAVQTAQNGVRAMSSLGSSG 441
Qy 309 LGGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQGLGLPV 368
Db 442 LGGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQGLGLPV 501
Qy 369 GQMGARAGGSLGVLRVPPPYVMPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 428
Db 502 GQMGARAGGSLGVLRVPPPYVMPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 561
Qy 429 GPQVNTKLGYNNAVAGCTGIVDPNGVLTNNHVIAGATDINAFSGSGGTGVDVV 488
Db 562 GPQVNTKLGYNNAVAGCTGIVDPNGVLTNNHVIAGATDINAFSGSGGTGVDVV 521
Qy 489 GYDRTQDVAVLQLRGAGGLPSAAGGVAVGEPVWAMGSGGGTTPRAVPGRVVAGT 548
Db 622 GYDRTQDVAVLQLRGAGGLPSAAGGVAVGEPVWAMGSGGGTTPRAVPGRVVAGT 681
Qy 549 VQASDSLTAETLNGLIQFDAAIQPDAGGPPVNGLGQVGMNTAAS 596
Db 682 VQASDSLTAETLNGLIQFDAAIQPDAGGPPVNGLGQVGMNTAAS 729

RESULT 15
AAE17573
ID AAE17573 standard; protein; 729 AA.
AC AAE17573;
XX
XX 22-APR-2002 (first entry)
XX
XX Mycobacterium species MTB72FmutSA fusion protein.
XX
XX Fusion protein; antigen; serological sensitivity; immune response;
XX tuberculosis; infection; vaccine; MTB72FmutSA; Ra12-TbH9-Ra35MutSA;
XX mutant; mutein.
XX
XX Mycobacterium sp.
XX
XX Key Location/Qualifiers
XX Region 8..139
XX /label= Ra12_protein
XX Region 142..532
XX /label= TbH9PL_protein
XX Region 535..729
XX /label= Ra35_protein
XX Misc-difference 710
XX /note= "Wild type Ser substituted with Ala"
XX
XX WO200198460-A2.
XX
XX 27-DEC-2001.
XX
XX 20-JUN-2001; 2001WO-US019959.
XX
XX 20-JUN-2000; 2000US-00597796.
XX
XX 01-FEB-2001; 2001US-0265737P.
XX
XX (CORI-) CORIXA CORP.
XX
XX Skeiky Y, Reed S, Alderson M;
XX
XX WPI; 2002-147798/19.
XX
XX DR N-PSDB; AAD28343.
XX
XX Composition comprising MTB39 antigen and MTB32A antigen from
XX Mycobacterium species, useful for eliciting immune response in a subject.
XX
XX Claim 88; Fig 5; 136pp; English.
XX
XX
```

```
XX The present invention relates to fusion proteins containing at least two
CC Mycobacterium species antigens, nucleotides encoding them and
CC compositions comprising such fusion proteins. The present invention
CC particularly relates to nucleic acids encoding fusion proteins that
CC include two or more individual M. tuberculosis antigens which increase
CC the serological sensitivity of sera from individuals infected with
CC tuberculosis and methods for their use in diagnosis, prevention and
CC treatment of tuberculosis infection. Sequences of the invention are
CC useful for eliciting an immune response in a mammal, e.g., human,
CC immunised with BCG. They are useful in the diagnosis, treatment and
CC prevention of Mycobacterium infection. The fusion proteins and the
CC polynucleotides are useful as diagnostic tools in patients infected with
CC Mycobacterium, in vitro and in vivo assays for detecting humoral
CC antibodies or cell-mediated immunity against M. tuberculosis, for the
CC diagnosis of an infection or monitoring of disease progression, as
CC immunogens to generate or elicit a protective immune response in a
CC patient and for raising anti-M. tuberculosis antibodies in a non-human
CC animal. Sequences of the invention are also used as vaccines. MTB32A
CC fusion proteins of the invention are useful as in vivo diagnostic agents
CC for intradermal skin test. The present sequence is Mycobacterium species
CC MTB72FmutSA (Ra12-TbH9-Ra35MutSA) mutant fusion protein
XX
XX Sequence 729 AA;
XX
XX Query Match 97.9%; Score 2943; DB 5; Length 729;
XX Best Local Similarity 99.8%; Pred. No. 5.4e-194;
XX Matches 587; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 9 MVDGALPPEINSAEMVAGPGSASIVAAQAQWDSVADLPSAASAFQSVVGLTVGSWIG 68
Db 142 MVDGALPPEINSAEMVAGPGSASIVAAQAQWDSVADLPSAASAFQSVVGLTVGSWIG 201
Qy 69 SSAGLMVAASPFYVAMSVTAGQAELETAQAQVRAAAAYETAYGLTVPPIAENRAELMI 128
Db 202 SSAGLMVAASPFYVAMSVTAGQAELETAQAQVRAAAAYETAYGLTVPPIAENRAELMI 261
Qy 129 LIATNLIGQNTPATAVNEAEYGEWQAQDAAMFGYAAATATATATLPPPEAPEMTSAGG 188
Db 262 LIATNLIGQNTPATAVNEAEYGEWQAQDAAMFGYAAATATATATLPPPEAPEMTSAGG 321
Qy 189 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 248
Db 322 LLEQAAVEEASDTAAANQLMNNVPQALQQAQPTQGTTPSSKLGGLWKTVPSPHRSPI 381
Qy 249 MYSMANNHMSMTNSGVSMNTLSMLKGFAPAPAAAQAVQTAQNGVRAMSSLGSSG 308
Db 382 MYSMANNHMSMTNSGVSMNTLSMLKGFAPAPAAAQAVQTAQNGVRAMSSLGSSG 441
Qy 309 LGGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQGLGLPV 368
Db 442 LGGGVAANLGRAASVGSLSVPQAAANQAVTPAARALPLTSLTSAERGPQGLGLPV 501
Qy 369 GQMGARAGGSLGVLRVPPPYVMPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 428
Db 502 GQMGARAGGSLGVLRVPPPYVMPHSPAAAGDIAPPALSDQRFADFPALPLDPSAMVAQV 561
Qy 429 GPQVNTKLGYNNAVAGCTGIVDPNGVLTNNHVIAGATDINAFSGSGGTGVDVV 488
Db 562 GPQVNTKLGYNNAVAGCTGIVDPNGVLTNNHVIAGATDINAFSGSGGTGVDVV 621
Qy 489 GYDRTQDVAVLQLRGAGGLPSAAGGVAVGEPVWAMGSGGGTTPRAVPGRVVAGT 548
Db 622 GYDRTQDVAVLQLRGAGGLPSAAGGVAVGEPVWAMGSGGGTTPRAVPGRVVAGT 681
Qy 549 VQASDSLTAETLNGLIQFDAAIQPDAGGPPVNGLGQVGMNTAAS 596
Db 682 VQASDSLTAETLNGLIQFDAAIQPDAGGPPVNGLGQVGMNTAAS 729

Search completed: June 30, 2004, 16:48:49
JOB time : 63.2853 secs
```

